An Analytic Model of Angular Momentum Transport by Gravitational Torques: From Galaxies to Massive Black Holes

Philip F. Hopkins[1]∗, & Eliot Quataert[1]

[1]*Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720*

ABSTRACT

We present analytic calculations of angular momentum transport and gas inflow in galaxies, from scales of $\sim$ kpc to deep inside the potential of a central massive black hole (BH). We compare these analytic calculations to numerical simulations and use them to develop a sub-grid model of BH growth that can be incorporated into semi-analytic calculations or cosmological simulations. Motivated by both analytic calculations and simulations of gas inflow in galactic nuclei, we argue that the strongest torque on gas arises when non-axisymmetric perturbations to the stellar gravitational potential produces orbit crossings and shocks in the gas. This is true both at large radii $\sim 0.01 - 1$ kpc, where bar-like stellar modes dominate the non-axisymmetric potential, and at smaller radii $\lesssim 10$ pc, where a lopsided/eccentric stellar disk dominates. The traditional orbit crossing criterion is not always adequate to predict the locations of, and inflow due to, shocks in gas+stellar disks with finite sound speeds. We derive a modified criterion that predicts the presence of shocks in stellar dominated systems even absent formal orbit crossing. We then derive analytic expressions for the loss of angular momentum and the resulting gas inflow rates in the presence of such shocks. We test our analytic predictions using hydrodynamic simulations at a range of galactic scales, and show that they successfully predict the mass inflow rates and quasi-steady gas surface densities with a small scatter $\simeq 0.3$ dex. We use our analytic results to construct a new estimate of the BH accretion rate given galaxy properties at larger radii, for use in galaxy and cosmological simulations and semi-analytic models. While highly simplified, this accretion rate predictor captures the key scalings in the numerical simulations. By contrast, alternate estimates such as the local viscous accretion rate or the spherical Bondi rate fail systematically to reproduce the simulations and have significantly larger scatter.

Key words: quasars: general — galaxies: active — galaxies: evolution — cosmology: theory

1 INTRODUCTION

Gas inflows into galactic nuclei play a critical role in galaxy formation, initiating a wide range of phenomena including starbursts, the formation of bulges, spheroidal galaxies, and nuclear star clusters, and accretion onto supermassive black holes (BHs). The initial triggers for these inflows are diverse and may include galaxy mergers and tidal encounters; bar and spiral disk instabilities; secondary and tertiary instabilities ("bars within bars"); and large-scale triaxial or binary structures in galaxy halos.

Intense star formation in galactic nuclei also has a major effect on the structure of the resulting systems, with dissipative processes dominating the formation of the inner $\sim$kpc of bulges at surface mass densities reaching $\sim 10^{11-12} M_\odot$ kpc^{-2} (Ostriker 1980; Carlberg 1986; Gunn 1987; Kormendy 1989; Hernquist et al. 1993;

Hopkins et al. 2009a,c). There is also strong indirect evidence that the most luminous starbursts, ultra-luminous infrared galaxies (ULIRGs), are accompanied or followed by periods of luminous quasar activity (Sanders et al. 1988a,b; Dasyra et al. 2007).

On galactic scales, numerical simulations have shown that in gas-rich mergers, tidal torques lead to rapid gas inflow into the central $\lesssim$ kpc (Hernquist 1989; Barnes & Hernquist 1991, 1996). Minor mergers and instabilities in self-gravitating disks (bars and spiral waves) can drive similar inflows (Hernquist & Mihos 1995; Bournaud et al. 2005; Eliche-Moral et al. 2008; Younger et al. 2008). And in high redshift, gas-rich disk galaxies, the sinking of massive star-forming clumps may produce analogous strong gravitational torques (Bournaud et al. 2007; Genzel et al. 2008).

How gas is transported from ~ 1 kpc to the much smaller scales relevant for massive BH growth remains uncertain (e.g., Goodman 2003). Once gas reaches sub-kpc scales, the torques from large-scale bars, and even from major mergers, become less ef-

∗ E-mail:phopkins@astro.berkeley.edu

ficient. Local viscous stresses – which are believed to dominate angular momentum transport near the central BH (e.g., Balbus & Hawley 1998) – are inefficient at radii $\gtrsim 0.01 - 0.1$ pc (e.g., Shlosman & Begelman 1989; Goodman 2003; Thompson et al. 2005). And scattering of individual gas clumps or molecular clouds cannot produce the high accretion rates needed to explain bright quasars (Hopkins & Hernquist 2006; Kawakatu & Wada 2008; Nayakshin & King 2007). As a consequence, most models implicitly or explicitly invoke some form of global gravitational torques ("bars within bars"; Shlosman et al. 1989) to continue transport to smaller radii. Recently, numerical simulations using a variety of techniques have been able to follow the inflow of gas to smaller scales in galaxies (Escala et al. 2004; Escala 2007; Mayer et al. 2007; Wise et al. 2008; Levine et al. 2008; Hopkins & Quataert 2010a). This has allowed simulations to begin to address some of the long-standing questions about gas inflow in galactic nuclei.

In a previous paper (Hopkins & Quataert 2010a), we presented a large suite of ~ 100 simulations that begin from initial conditions of either gas-rich major mergers or bar-unstable disks, and follow the inflow of gas from > 10 kpc to < 0.1 pc; at the smallest radii, the system begins to resemble a traditional, viscous accretion disk. Our method uses a series of independent numerical simulations on successively smaller scales (rather than an AMR or particle-splitting approach), which allows us to systematically survey how a variety of initial conditions and galaxy properties affect gas dynamics in galactic nuclei. These calculations show that large-scale bars can indeed drive gas to $\lesssim$ kpc; and if the disky material at smaller radii becomes self-gravitating, it triggers a cascade of secondary gravitational instabilities that transports gas deeper in the potential well. This qualitatively resembles the "bars within bars" scenario. However, there are several critical differences. First, the secondary instabilities on intermediate scales ($\sim 10 - 100$ pc) exhibit a diverse range of morphologies. A more accurate characterization would thus be "it's non-axisymmetric features all the way down" (or "stuff within stuff"). This is consistent with the fact that observed systems that do show nuclear asymmetries can show nuclear rings, barred rings, one or three-armed modes, and clumpy/irregular structures (Martini & Pogge 1999; Peletier et al. 1999; Knapen et al. 2000; Laine et al. 2002; Knapen et al. 2002; Greene et al. 2009). Moreover, in our simulations, the inflow rates and gas morphology are time-variable, leading to modest duty cycles that may explain the generally weak correlation in observations between mode structure and inflow/outflow on different spatial scales. A second key difference between our numerical results and the canonical "bars within bars" model is that once gas reaches the radii where the gravity of the BH begins to dominate (~ 10 pc), bar-like modes no longer survive. However, a new instability appears, in the form of a slowly precessing lopsided/eccentric disk or one-armed spiral ($m = 1$) mode, which dominates the angular momentum transport to < 0.1 pc. The resulting torques can drive gas inflows at rates comparable to those required to sustain the brightest observed quasars, $\sim 10 M_\odot$ yr^{-1}. And the surviving stellar relics of these modes closely resemble the observed nuclear disks in M31 (Lauer et al. 1993) and other galaxies (Hopkins & Quataert 2010b).

Given the simplifying assumptions necessary in numerical simulations (e.g., sub-resolution star formation and feedback models), it is clearly important to understand analytically the physics of angular momentum transport. Moreover it is not possible to simulate the entire parameter space relevant to galactic inflows, so a way of generalizing the numerical results would be useful, both for gaining physical intuition and for developing subgrid models that

can be used in calculations that lack the physics/resolution to directly capture gas inflow on sub-kpc scales.

There is at present no analytic theory that accurately predicts the inflow rates of gas in mixed stellar-plus-gas systems, despite the rich history of analytic studies of gravitational instabilities in disks (see e.g. Goldreich & Lynden-Bell 1965a,b; Lin et al. 1969; Toomre 1969; Lynden-Bell & Kalnajs 1972; Goldreich & Tremaine 1980; Binney & Tremaine 1987; Shu et al. 1990; Toomre 1977, 1981; Lau & Bertin 1978; Adams et al. 1989; Shu et al. 1990; Ostriker et al. 1992). Almost all previous calculations have simplified the problem dramatically by focusing on *strictly* stellar or gas disks (as opposed to gas+stellar disks) and by ignoring the presence of shocks and cooling/dissipation in the gas. In addition, the majority of the analytic development and intuition has focused on the case of weak perturbations. In these limits, the leading-order instantaneous gravitational torques associated with the asymmetric gravitational potential cancel exactly. Gravitational angular momentum exchange occurs in narrow resonances (Goldreich & Tremaine 1979), and the net rate at which the gas loses angular momentum is second-order in the fractional magnitude of the perturbation (denoted $|a|$ throughout; Lynden-Bell & Kalnajs 1972). It is typically further suppressed by strong powers in the radial and azimuthal wavenumbers of the perturbation (in both the global and local perturbation limits; Kalnajs 1971). These results lead to the prediction of relatively weak torques, with sub-sonic mean gas inflow velocities $v_{\rm inflow} \lesssim 0.1\,|a|^2\,Q^{-1}\,c_s$ (where Q is the standard Toomre Q parameter and $c_s \ll V_c$ is the gas sound speed; Goodman 2003).

In the numerical simulations described above of galaxy mergers, disk instabilities, and secondary and tertiary instabilities, none of the simplifications often utilized in the study of angular momentum transport are in fact realized. Most of the angular momentum exchange occurs in the non-linear regime in which the non-axisymmetric perturbations to the gravitational potential due to the stars are sufficiently large that they induce orbit crossing, shocks, and dissipation in the gas. The multi-component dissipative nature of the system is thus critical. In, for example, instabilities of galactic disks, the gas orbits respond to the asymmetric potential of the collisionless (stars plus dark matter) material; when the asymmetry is sufficiently large, the gas is driven into shocks, which dissipates energy and allows the torques from the asymmetric potential to remove angular momentum efficiently. The "pile up" of material in these shocks manifests itself observationally as dust lanes and enhanced star formation in the leading edge of spiral arms (Roberts et al. 1979; Athanassoula 1992b; Wada 1994; Bournaud et al. 2005). The same basic process also dominates the angular momentum loss in major galaxy mergers, with the asymmetry in the stellar potential being excited by the companion (Hernquist 1989; Barnes & Hernquist 1991, 1996; Hopkins et al. 2009b). The resulting mean inflow velocities are typically super-sonic, with inflow rates up to $\sim 100 - 1000$ times larger than the predictions based on linear torques in single-component systems! The simulations also explicitly demonstrate that "hydrodynamic" pressure forces do not contribute significantly to the torques and inflow (see Barnes & Hernquist 1996; Berentzen et al. 2007; Hopkins et al. 2009b; Hopkins & Quataert 2010a): the process driving angular momentum loss is overwhelmingly gravitational.

The basic processes summarized above – gas responding to gravitational torques predominantly from stars, "piling up" in shocks, and exhibiting supersonic mean radial inflow/outflow velocities as a result – have been directly observed in a variety of nearby galaxies, where the role of gravitational torques can be mapped (Haan et al. 2009; Durbala et al. 2009; Block et al. 2001;

Sakamoto et al. 1999; Quillen et al. 1995). These observations also extend to galactic nuclei where secondary instabilities ("bars within bars") may become important (Laine et al. 2002; Schinnerer et al. 2006; Knapen et al. 2002), and to the gas surrounding bright AGN (Combes et al. 2004; García-Burillo et al. 2005). The torques inferred in many of these systems correspond to the gas losing a large fraction of its angular momentum in a few dynamical times. This is much stronger than can be accounted for by torques in linear, single-component systems and highlights the need for a better physical understanding of gravitational torques and inflow under realistic galactic conditions.

In this paper, we develop an analytic model that attempts to account for the realistic complications of angular momentum redistribution in systems with both collisional (gas) and collisionless (stellar/dark matter) components, in the presence of shocks and dissipation. This is an analytic companion to the numerical simulations in Hopkins & Quataert (2010a), which in turn builds on the large history of analytic and numerical studies of angular momentum transport in galaxies described above. The remainder of this paper is organized as follows. In § 2, we calculate the response of gas to general asymmetric perturbations, deriving the conditions for shocks and the resulting angular momentum transport; we consider the linear regime of no shocks (§ 2.2), strong orbit crossings that induce shocks (§ 2.3), and marginal orbit crossings (§ 2.3.3). We then briefly describe the balance between inflow and star formation in gaseous disks, and the resulting quasi-steady gas density distribution (§ 3). In § 4, we describe our suite of high-resolution hydrodynamic simulations of mergers and unstable disks, with multiscale re-simulations of secondary instabilities, which we use to test our analytic scalings as a function of galaxy properties and physical scale. In § 5, we use the results of § 2 to develop an analytic model that predicts the inflow rate to the BH given the conditions at large radii in a galaxy; we show that this predictor is much more accurate than the spherical accretion prediction or a local viscous model. Finally, we summarize and discuss our results in § 6.

2 GAS RESPONSE IN ASYMMETRIC SYSTEMS

2.1 Definitions

We adopt a cylindrical coordinate frame, (R, ϕ, z), and consider an initially axisymmetric disk with an arbitrary spherical (BH+bulge+halo) component. We use the test particle approximation to calculate the response of the disk to a specified nonaxisymmetric perturbation to the potential, largely neglecting pressure forces. Since we are interested in the behavior of cold, rotationally supported, gas, the gas of interest is always in a thin ($h/R \ll 1$) disk, aligned with the stellar disk.

The initial potential in the disk plane can be written $\Phi_0 = \Phi_0(R)$, and other properties are defined using standard notation:

$$\Phi_0 = \Phi_0(R) \tag{1}$$

$$V_c^2 = R \frac{\partial \Phi}{\partial R} \approx \frac{GM_{\mathrm{enc}}(<R)}{R} \tag{2}$$

$$\Omega \equiv t_{\mathrm{dyn}}^{-1} = V_c/R \tag{3}$$

$$\kappa^2 \equiv R \frac{d\Omega^2}{dR} + 4\Omega^2 = \frac{\partial^2 \Phi}{\partial R^2} + 3\Omega^2 \tag{4}$$

where V_c is the circular velocity, Ω the angular velocity, and κ the epicyclic frequency. We now consider a global perturbation in the disk plane and define the perturbed potential

$$\Phi \rightarrow \Phi_0(R) + \Phi_1(R, \phi) \tag{5}$$

In what follows, the subscript "1" will denote all perturbed quantities. We will consider a frame rotating with the perturbation pattern speed Ω_p, so that in these coordinates the perturbation is stationary. An unperturbed orbit in the rotating frame is at fixed R with angular speed $\dot{\phi}_0 = \Omega - \Omega_p$. We can represent the perturbed potential as a sum over modes:

$$\Phi_1 \equiv a(R)\, \Phi_0\, f(\phi) = \Phi_0(R) \sum_{m=1}^{\infty} a_m(R) \cos(m\phi)\,. \tag{6}$$

For all perturbed quantities X, we define $X_1(R, \phi) = X_a(R) \cos(m\phi)$. Thus $\Phi_a = |a|\, \Phi_0$, etc. Note that for a given particle, the perturbed and unperturbed quantities depend implicitly on time via $R(t)$ and $\phi(t)$.

2.2 Torques and Inflow: Linear Regime

When the perturbation is weak ($|a| \ll 1$), it is straightforward to expand the equations of motion for a test particle in the perturbed potential using the epicyclic approximation. With the substitutions

$$R \rightarrow R_0 + R_1(R, t) = R_0 + R_a \cos[m(\Omega - \Omega_p)t]$$
$$\phi \rightarrow \phi_0(t) + \phi_1(R, t) = \phi_0(t) + \phi_a \cos[m(\Omega - \Omega_p)t] \tag{7}$$

the linearized equations of motion become:

$$R_1(R, t) = \left(\frac{f_0}{\Delta}\right) \cos[m(\Omega - \Omega_p)t]$$
$$\dot{\phi}_1(R, t) = -2\Omega \frac{R_1(t)}{R_0} - \frac{\Phi_a}{R^2(\Omega - \Omega_p)} \cos[m(\Omega - \Omega_p)t] \tag{8}$$

where

$$f_0 \equiv -\left[\frac{d\Phi_a}{dR} + \frac{2\Omega \Phi_a}{R(\Omega - \Omega_p)}\right]_R \tag{9}$$

$$\Delta \equiv \kappa^2 - m^2(\Omega - \Omega_p)^2\,. \tag{10}$$

When the orbits of particles do not cross, i.e., there are no shocks, the net torque integrated over an orbit is identically zero to first order in Φ_1. To see this, note that the instantaneous torque per unit mass τ is

$$\tau \equiv -(\mathbf{R} \times \nabla\Phi)_z = -\frac{\partial \Phi}{\partial \phi} \tag{11}$$

Integrated over an orbit, the net angular momentum loss is therefore

$$\Delta J = \int \tau\, dt$$
$$= -\int_0^{2\pi} \frac{\partial \Phi}{\partial \phi} (\dot{\phi})^{-1}\, d\phi$$
$$= -\int_0^{2\pi} \Phi_a(R) \frac{\partial f(\phi)}{\partial \phi} (\dot{\phi})^{-1}\, d\phi\,. \tag{12}$$

To first order, we must use $\dot{\phi} = \dot{\phi}_0 = \Omega - \Omega_p$, so that $\Delta J \approx -\Phi_a(R)(\Omega - \Omega_p)^{-1} \int_0^{2\pi} (\partial f(\phi)/\partial \phi) d\phi$, which trivially evaluates to zero because $f(\phi)$ is periodic.

Expanding the orbits to second order does not change this conclusion – the net torque is again zero. Goldreich & Tremaine (1978, 1979) and Athanassoula (1992a, 2003) derive this result rigorously. They show that there can only be a non-zero second-order torque at the resonance points – where the epicyclic expansion breaks down – or if the perturbation magnitude is growing in time at a rate that is large compared to $\Omega - \Omega_p$.

The above derivation applies for collisionless particles. However, gas is collisional, and thus obeys a slightly different dispersion relation. For systems consisting of both gas and stars – which

are the focus of this paper – the gas mode is offset from the stellar mode by a small angle δ due to dissipation. This gives rise to a torque on the gas by the stars (typically $\delta \sim 10 - 15$ deg in simulations; Noguchi 1988; Barnes & Hernquist 1996; Barnes 1998; Berentzen et al. 2007). Following Wada (1994), the phase shift δ and the resulting orbit-averaged torque on the gas can be derived in the limit of weak modes that do not induce orbit crossing or shocks in the gas:

$$\tau = \frac{dJ}{dt} = -\frac{m}{2} f_* V_c^2 \left| \frac{\Phi_a(R)}{V_c^2} \right|^2 \sin(\delta) \tilde{B}(R) \qquad (13)$$

where f_* is the stellar mass fraction in the disk,

$$\delta = \tan^{-1} \left(\frac{m \kappa (\Omega_p - \Omega) \Lambda}{\Delta} \right), \qquad (14)$$

$$\tilde{B} = \frac{(2[1 - \Omega_p/\Omega]^{-1} + \partial \ln \Phi_a/\partial \ln R)^2}{\Omega^{-2} \sqrt{\Delta^2 + m^2 \Lambda^2 \kappa^2 (\Omega - \Omega_p)^2}}, \qquad (15)$$

and $\Lambda \sim 0.1$ is a dimensionless constant that parameterizes the dissipation in the gas (formally, Λ is the dimensionless constant in a phenomenological 'friction' term in the momentum equation that is proportional to the radial velocity; the value ~ 0.1 is motivated by comparison with hydrodynamic simulations). Note that the net torque is second-order in the perturbation amplitude $|a| \sim \Phi_a/V_c^2$. In addition, for a bar ($m = 2$) mode inside co-rotation, $\delta > 0$, and the gas mode leads the stellar mode (implying net angular momentum loss by the gas); this reverses outside co-rotation.

Since the gas follows nearly circular orbits ($J \approx \Omega R^2$), the orbit-averaged gas inflow rate follows trivially from the torque $\tau = dJ/dt$:

$$\dot{M} = 2\pi \Sigma_{\text{gas}} R^2 \Omega \frac{\tau}{V_c^2 (1 + \partial \ln V_c/\partial \ln R)} . \qquad (16)$$

Note the sign convention: positive $\dot{M}$ is outflow, negative is inflow. For the simple case of a constant-V_c disk with $\Phi_a \sim |a| V_c^2$ and typical values for $\delta \sim 10°$ calibrated from simulations, equation 13 implies

$$\dot{M} \approx -|a|^2 \frac{\alpha m \Sigma_{\text{gas}} R^2 \Omega}{(1 - \Omega_p/\Omega)^2 |2 - m^2 (1 - \Omega_p/\Omega)^2|} \qquad (17)$$

where $\alpha \sim 2 - 3$. Far from resonance, this reduces to $\dot{M} \sim |a|^2 M_{\text{gas}} \Omega$.

In addition to the torque between the gas and stars, non-axisymmetric waves can also directly transport angular momentum (with the coupling at resonances); this yields a qualitatively similar torque that is also second-order in $|a|$ (Lynden-Bell & Kalnajs 1972). Kalnajs (1971) derives an exact solution for the logarithmic spiral, for which the inflow rate can be written

$$\dot{M}_{\text{K71}} = -|a|^2 \left(\frac{c_s}{V_c} \right) \frac{\partial |K|^2}{\partial |kR|} \frac{\pi^2 \sqrt{3} m}{2\sqrt{2} Q} \Sigma_{\text{gas}} R^2 \Omega \qquad (18)$$

where $K \equiv \Gamma[(m + ikR + 1/2)/2]/\Gamma[(m + ikR + 3/2)/2]$ and k is the radial wavenumber of the mode. The resulting torque scales as $|a|^2$ but is smaller than the torque between gas and stars (eq. 13) by two factors: (1) a term $\sim (c_s/V_c) Q^{-1} \sim \Sigma_{\text{gas}}/\Sigma_{\text{tot}}$, and (2) a pre-factor $\partial |K|^2/\partial |kR|$ that depends on the structure of the mode, which scales as $\approx |kR|/(1 + m^3)$ for $|kR| \ll 1$ or $\sim |kR|^{-2}$ for $|kR| \gg 1$. As a result, the predicted torques become small in both the global and local mode limits.

2.3 Torques & Inflow: Non-Linear Regime (Orbit Crossing)

2.3.1 General Criteria

In the epicyclic approximation of the previous subsection, the perturbation to the orbit of a particle is periodic in ϕ. For a gaseous disk, however, this assumption breaks down if the perturbed orbits of particles cross each other, because orbit crossing in gas leads to dissipation (generally shocks) and inflow. Such dissipation breaks the periodicity in ϕ that leads to the cancellation of the first order (in $\Phi_a \propto a$) contribution to the torque in equation (12). As a result, in the limit of strong orbit crossing an order of magnitude estimate of the inflow rate is

$$\dot{M} \sim \frac{-|a| m \Sigma_{\text{gas}} R^2 \Omega}{1 + \partial \ln V_c/\partial \ln R} \sim -|a| \Sigma_{\text{gas}} R^2 \Omega. \qquad (19)$$

Equation 19 will derived more rigorously in what follows.

Papaloizou & Pringle (1977) derived a useful criterion for when orbits cross. Consider two quasi-circular orbits at the same ϕ, with infinitesimal initial radial separation δr_i. Using the linear solution for each particle (eq. 8), the separation between the particles as a function of time is given by $\delta r(t) = \delta r_i [1 + (\partial R_a/\partial R_0) \cos(m\phi)]$. Thus if

$$|\zeta| \equiv |\partial R_a/\partial R_0| = \left| \frac{1}{\Delta} \left(\frac{\partial f_0}{\partial R} - \frac{f_0}{\Delta} \frac{\partial \Delta}{\partial R} \right) \right| \gtrsim 1, \qquad (20)$$

then the orbits cross, i.e., $\delta r < 0$, at some point (phase) in the orbit. We caution that this is a sufficient, but not necessary criterion for orbit crossing. One can readily derive a similar criterion for when the particles cross in azimuth: $|\partial \phi_a/\partial \phi| > 1$; and, clearly, if $|R_a| > R$ there will be orbit crossing. Near-resonances, one can have orbit-trapping (where the gas librates about a specific radius), which will lead to shocks, for $|\partial R_a/\partial R|$ as small as $\sim 0.05 - 0.1$ (see Binney & Tremaine 1987). In general, even for linear modes, there is no purely local criterion for orbit crossing – one of the above conditions must be satisfied *somewhere* for orbit crossings to occur, but if so, it can in principle occur anywhere. A full census requires calculating the mode structure everywhere and the resulting particle orbits. Nevertheless we can make considerable analytic progress in several limits.

2.3.2 Strong Orbit-Crossing Limit

We now derive the net torque and inflow rate in the limit $|\zeta| \gg 1$, in which the gas experiences strong orbit crossings and shocks. Prior to equation 20, we derived the pericentric separation of two perturbed particles at a given ϕ. Generalizing this result, the separation after moving from some arbitrary ϕ_i to a final ϕ_f will be given by $\delta r(t) = \delta r_i[1 + (\partial R_a/\partial R_0)(f(m\phi_f) - f(m\phi_i))]$, where $R_1 = R_a f(m\phi)$ represents the epicyclic expansion solution, i.e., $f(m\phi) = \cos(m\phi)$. Because $\zeta \gg 1$, there is orbit crossing at a range of ϕ. We are thus not in a regime in which the epicyclic approximation is valid in a global sense; however, we can consider it locally valid for a small amount of time ("between shocks"). Globally, the gas orbits may be quite non-circular but we can approximate this motion with a series of small epicyclic 'steps' separated by shocks that dissipate angular momentum and energy, allowing the gas to flow in.

Given an initial separation of δr_i at ϕ_i, the orbits will cross at a final ϕ_f such that $f(m\phi_f) - f(m\phi_i) = -1/\zeta$. We assume that dissipation is efficient when the gas streams cross and shock because the cooling time of the gas is always short under the conditions of interest. As a result, the colliding streams will dissipate the kinetic energy of their relative bulk motion. This effectively "resets" the orbit to a circular orbit with the appropriate total angular

momentum per unit mass of the streams at that time. We use this approximation to estimate the angular momentum loss due to the non-axisymmetric torques; the same results could be derived using energy considerations.

The angular momentum per unit mass in the rest-frame is $j = R^2(\dot\phi + \Omega_p) = (R_0 + R_1)^2(\dot\phi_0 + \dot\phi_1 + \Omega_p)$ (recall that ϕ is defined in the frame rotating at Ω_p). Expanding to first order, the change in angular momentum of the streams from ϕ_i to ϕ_f is $\Delta j \equiv j_f - j_i \approx R_0(2R_a\Omega + R_0\dot\phi_a)(f[m\phi_f] - f[m\phi_i])$ where we have used $R_1 = R_a f(m\phi)$ and $\dot\phi_1 = \dot\phi_a f(m\phi)$. Using equation 8 and the condition for when the streams shock ($f[m\phi_f] - f[m\phi_i] = -1/\zeta$), we find

$$\Delta j = -\frac{R[2R_a(\dot\phi_0 + \Omega_p) + R_0\dot\phi_a]}{\zeta} \quad (21)$$

$$= \frac{\Phi_a}{\Omega - \Omega_p}\frac{1}{\zeta} \quad (22)$$

The time required to move in azimuth by $\Delta\phi = \phi_f - \phi_i$ is given by

$$\Delta t = \frac{\Delta\phi}{|\Omega - \Omega_p|} = \frac{f^{-1}(f(m\phi_i) - \zeta^{-1}) - m\phi_i}{m|\Omega - \Omega_p|} \quad (23)$$

Note that because the $f(m\phi)$ term ranges from -1 to $+1$, the sign of $\zeta = \partial R_a/\partial R_0$ in the above simply amounts to a phase offset in the shock location, so the above can be generalized by the replacement $\zeta \to |\zeta|$.

The average rate of angular momentum loss is simply given by $\Delta j/\Delta t$. Using the above solutions for $\Delta t(\phi_i)$ and Δj, this can be written as a sum over the "number of shocks" encountered in a circuit of $\Delta\phi = 2\pi$. In the $|\zeta| \gg 1$ limit, we write this as continuous average over ϕ, with $\Delta j/\Delta t \to \langle dj/dt \rangle = \tau$ and

$$\langle\tau\rangle = \frac{1}{2\pi}\int\frac{\Delta j}{\Delta t[\phi_i]}d\phi_i \quad (24)$$

where $\langle\tau\rangle$ is the azimuthally averaged torque. The general expression for τ cannot be written in closed form, but the leading-order term in ζ^{-1} can be used to derive the following expressions for the torque τ, inflow rate $\dot M$, and radial velocity V_R:

$$\langle\tau\rangle = \frac{m}{2\pi}\Phi_a F(\zeta)\,\text{sign}(\Omega - \Omega_p) \quad (25)$$

$$\dot M = (\Sigma_{\text{gas}}R^2\,\Omega)\left[\frac{\Phi_a}{V_c^2}\right]\frac{m\,\text{sign}(\Omega - \Omega_p)}{(1 + \partial\ln V_c/\partial\ln R)}F(\zeta) \quad (26)$$

$$\langle V_R\rangle = V_c\left[\frac{\Phi_a}{V_c^2}\right]\frac{m\,\text{sign}(\Omega - \Omega_p)}{(1 + \partial\ln V_c/\partial\ln R)}F(\zeta) \quad (27)$$

$$F(\zeta) \approx \left[2 - \frac{1}{2|\zeta|} + \mathcal{O}(|\zeta|^{-2})\right]\text{sign}(\zeta) \quad (|\zeta| \gg 1) \quad (28)$$

Note that the net torque and inflow rate are independent of ζ for $\zeta \gg 1$. They depend primarily on the magnitude of the non-axisymmetric potential Φ_a. This is because although the change in angular momentum per shock scales as ζ^{-1} (eq. 22), the time between shocks itself also scales as ζ^{-1} (this follows from Taylor expanding eq. 23 and averaging over ϕ_i). Another way to say this is that there are $N \sim \zeta$ shocks each of which resets the gas to a circular orbit with $\Delta j \propto 1/\zeta$, so that the net torque is independent of ζ.

Dimensionally, given $\Phi_a \sim |a|V_c^2$, equation 26 reduces to $\dot M \sim -|a|\Sigma_{\text{gas}}R^2\Omega$, in agreement with our order of magnitude estimate in equation 19. Thus in the strong-torque regime the inflow rates are *linear* in $|a|$: when orbit crossings are important, gas inflow can easily be enhanced by an order of magnitude or more relative to the expectations from the weak-torque theory of §2.2.

The sign() terms in equations 25-28 are important and arise from terms that in the full solution are continuous; i.e.,

$\zeta/\sqrt{1+\zeta^2} \to \text{sign}(\zeta)$ for $|\zeta| \gg 1$. Recall that with our sign convention, Φ_a is negative where Σ_a is positive. For typical parameter values inside of co-rotation ($\Omega > \Omega_p$), ζ is negative and therefore $\dot M$ is also negative (inflow), whereas outside of corotation it is positive (outflow). This dependence on corotation is essentially a question of whether the shocks are in prograde or retrograde motion in the pattern frame, which changes at the corotation resonance. The dependence on $\text{sign}(\zeta)$ is due to the phase between the shocks and the surface density perturbation: when $\zeta < 0$, shocks occur where $f(m\phi) - f(m\phi_i)$ is maximized – in phase with the positive maximum in Σ_1 – and hence where (for $\Omega > \Omega_p$) Δj is positive. This generically leads to outflow. But when $\zeta < 0$, the shocks occur leading the mode, where Δj is negative and minimized with respect to the unperturbed orbit; this leads to inflow.

If instead of using real variables, we generalize to complex Φ_a and ζ, a nearly identical derivation to that presented here leads to $\Phi_a\,\text{sign}(\zeta) \to |\Phi_a|\cos(\arg[\Phi_a\zeta^*])$, i.e., $|\Phi_a|\cos\tilde\phi$, where $\tilde\phi$ is the phase angle between ζ and Φ_a; this can have an in-phase inflow and/or out-of-phase outflow term. Likewise, the $\text{sign}(\Omega - \Omega_p)$ term can be generalized for complex $\omega = m\Omega_p + i\gamma$, by taking $\text{sign}(\Omega - \Omega_p) \to \text{Re}[m|\Omega - \Omega_p|/(m\Omega - \omega)]$.

2.3.3 Marginal Orbit-Crossing Limit

The derivation in §2.3.2 is only valid when $|\zeta| \gg 1$. Indeed, when $|\zeta| \lesssim 1$, the quantities used in these derivation are undefined, even though it is still very much possible that there are shocks and even orbit crossings (as discussed in § 2.3.1). Moreover, the requirement of formal orbit crossings for strong torques and shocks is a consequence of the fact that we have thus far (implicitly) considered only infinitely cold disks. There can, in fact, be shocks when $|\zeta| \lesssim 1$, even without orbit crossings, if the non-axisymmetric potential is dominated by the stars rather than the gas. We now estimate the resulting torques and inflow rate in this limit.

If the disk is not gas-dominated, then the gas orbits are given roughly by the standard equations of motion in response to the stellar potential. Given radial variations in some property of the gas, the length-scale h_{eff} at which pressure forces become important is given by $h_{\text{eff}} = 2c_{\text{eff}}^2/G\Sigma$ (e.g., Toomre 1964) where c_{eff} is the effective dispersion of the gas, including both the true thermal sound speed and any non-thermal turbulent motions: $c_{\text{eff}}^2 = c_s^2 + v_t^2$. Note that in a spherical potential, h_{eff} is different from the vertical scale height h by $\sim 2c_{\text{eff}}\Omega/G\Sigma \sim Q$, the Toomre Q parameter (defined using c_{eff} not c_s). The stellar motions do not explicitly appear in the effective sound speed here, because the key criteria is whether *gas* shocks; Σ does, however, represent the mass density of gas plus disk stars since both participate in the mode self-gravity.

If stellar streamlines approach to within a distance $\lesssim h_{\text{eff}}$, they remain collisionless and conserve energy. But if gas that is initially spatially separated is forced into a *supersonic* 'near-collision' within a distance $\lesssim h_{\text{eff}}$ by the potential of the stars, it will shock and radiate the difference in energy between the streamlines (conserving only their total momentum). The key point is that when the non-axisymmetric potential is stellar dominated, gas "collisions" can happen not just at strict orbit crossings, but at a finite separation $\sim h_{\text{eff}}$ between streamlines.

If two gas streamlines approach one another at a relative velocity that is much smaller than the sound speed c_s, pressure effects can readjust the orbits without much dissipation. Thus significant dissipation probably requires relative stream velocities $> c_s$. What are the relative velocities? For radial motions, the maximum magnitude of the relative velocity over an orbit, from the epicyclic solution, is $\Delta v = |\partial v_{1,R}/\partial R|\,\delta r_i = [h_{\text{eff}}/(1 - |\zeta|)]|\partial\dot R_1/\partial R|$, where

in the second expression we have used the maximum initial separation streamlines can have such that they are separated by $h_{\rm eff}$ at some point in the orbit, $\delta r_i = h_{\rm eff}/(1-|\zeta|)$. The resulting criterion for gas streams to approach within a distance $\sim h_{\rm eff}$ with supersonic relative velocities $\Delta v > c_s$ is given by

$$\left| \frac{1}{\kappa} \frac{\partial v_{1,R}}{\partial R} \right| \left(\frac{c_{\rm eff}}{c_s} \right) \gtrsim \frac{1-|\zeta|}{2\pi Q}. \quad (29)$$

Since $v_{1,R}$ is linear in the mode amplitude $|a|$, we can define $|\partial v_{1,R}/\partial R|_{a=1}$ as the magnitude of $\partial v_{1,R}/\partial R$ for $|a|=1$, and write equation 29 as a condition on the mode amplitude

$$|a| \gtrsim \frac{(1-|\zeta|)}{2\pi Q} \left(\frac{c_s}{c_{\rm eff}} \right) \left[\kappa^{-2} \left(\left| \frac{\partial v_{1,R}}{\partial R} \right|_{a=1}^2 + \left| \frac{\partial v_{1,\phi}}{\partial R} \right|_{a=1}^2 \right) \right]^{-1/2} \quad (30)$$

where we have also included the relative azimuthal velocity $v_{1,\phi}$ for completeness.

In equations 29 and 30 we have allowed for the possibility that the disk has non-thermal random velocities via the effective sound speed $c_{\rm eff}$, which is also in our definition of Q. One subtlety is whether the condition for effectively dissipating the non-axisymmetric motions is a relative velocity $\Delta v > c_{\rm eff}$ or $\Delta v > c_s$. We suspect that the latter is appropriate. When $c_{\rm eff} > \Delta v > c_s$, the energy per unit mass dissipated via non-axisymmetric motions is small compared to the overall rate at which supersonic turbulence is dissipating in the ISM (the latter being due to, e.g., star formation). Nonetheless, the ordered velocities induced by the non-axisymmetric potential can be efficiently dissipated if $\Delta v > c_s$; we confirm this below in simulations that have resolved turbulent $c_{\rm eff} > c_s$.

Equations 29 and 30 cannot be satisfied for $|a| < 1$ in a sufficiently cold disk with $Q \ll 1$ and $|\zeta| \lesssim 1$. In this limit, the condition for shocks and orbit crossing reduces to our previous criterion, $|\zeta| > 1$. However, for larger Q, the effective "thickness" of a given gas stream increases and so for the same non-axisymmetric perturbation, it is easier to generate collisions of streamlines. Of course, for sufficiently large Q our assumption that the gas orbits are well described by test particles responding to the stellar potential will break down. But for the particularly relevant case of $Q \gtrsim 1$ and $c_{\rm eff} > c_s$, equations 29 and 30 show that shocks induced in the gas by the stellar potential are possible even though $|\zeta| \lesssim 1$. For typical modes (say, $m = 1$ modes in Keplerian potentials or bars in constant-V_c disks), at the order of magnitude level, this condition on the mode amplitude becomes $|a| \gtrsim 0.2\,Q^{-1}\,(c_s/c_{\rm eff})$. We reiterate that this is only true when the gravity of the perturbation is dominated by stars. For gas dominated systems, the mode structure would already include pressure effects, invalidating our analysis. The exact gas fraction at which this transition occurs is sensitive to the gas equation of state and phase structure; it is thus difficult to estimate precisely, beyond the order of magnitude estimate $M_{\rm gas} \gtrsim M_*$.

We now generalize the derivation in §2.3.2 to determine the inflow rate and torques on the gas in the $|\zeta| \lesssim 1$ limit. We assume that both ζ and $h_{\rm eff}/R$ are small, so that we can Taylor expand in both quantities. The separation between two streamlines is given by $\delta r(\phi) = \delta r_i (1 + \zeta [f(m\phi) - f(m\phi_i)])$ Gas that is initially within a maximum separation $\delta r_i = \delta R_{\max} \equiv h_{\rm eff}/(1-|\zeta|) \simeq h_{\rm eff}(1+|\zeta|)$ will collide at some point in the orbit. However, streamlines with initial separations less than a minimum value $\delta R_{\min}$ are never differentially accelerated by the perturbed potential to $\Delta v \sim c_s$ and thus never experience shocks. The exact value of $\delta R_{\min}$ depends on $|a|$, ζ, $|\partial v_{1,R}/\partial R|_{a=1}$, etc. As a rough estimate,

we take $\delta R_{\min} \sim h_{\rm eff}$ because gas initially separated by $\delta R_{\min} \lesssim h_{\rm eff}$ is strongly affected by internal pressure forces and is unlikely to be accelerated to relative velocities $\gtrsim c_s$. For a given δr_i satisfying $\delta R_{\min} < \delta r_i < \delta R_{\max}$, the streamlines reach $\delta r < h_{\rm eff}$ (i.e., 'collide') when $f(m\phi) = f(m\phi_i) + \zeta^{-1}(h_{\rm eff}/\delta r_i - 1)$. For $\delta r_i = R_{\min}$, this corresponds to $f(m\phi) = f(m\phi_i)$ and $\Delta\phi = 0$, while for $\delta r_i = R_{\max}$, the shocks occur at $f(m\phi) = f(m\phi_i) \pm 1$ and $\Delta\phi = \pm(\pi/2m)$. When the gas reaches the ϕ such that $\delta r(\phi) < h_{\rm eff}$, it dissipates its relative orbital energy but conserves momentum. From §2.3.2, the change in specific angular momentum Δj for streamlines satisfying $\delta R_{\min} < \delta r_i < \delta R_{\max}$ is given by

$$\Delta j = -\frac{\Phi_a}{\Omega - \Omega_p} \frac{1}{\zeta} \left(\frac{h_{\rm eff}}{\delta r_i} - 1 \right). \quad (31)$$

The average change in angular momentum over all initial streamlines separations is given by

$$\langle \Delta j \rangle \simeq \frac{1}{h_{\rm eff}|\zeta|} \int_{h_{\rm eff}}^{h_{\rm eff}(1+|\zeta|)} \Delta j \, d(\delta r_i) \simeq \frac{1}{2} \frac{\Phi_a}{\Omega - \Omega_p} \, {\rm sign}(\zeta) \quad (32)$$

where we have used $R_{\min} = h_{\rm eff}$ and $R_{\max} = h_{\rm eff}(1 + |\zeta|)$ and in the second equality have Taylor expanded for $\zeta \ll 1$. Note that $h_{\rm eff}$ has cancelled out in our result for $\langle \Delta j \rangle$ – the value of the stream thickness does not influence how much angular momentum is lost.

The integral over δr_i in equation 32 is equivalent to an integral over the ϕ of the shocks. Thus the average over δr_i is equivalent to an average over ϕ and/or time. In particular, the range $\delta r_i = R_{\min}$ to $R_{\max}$ corresponds to $f(m\phi) = 0$ to ± 1. This is equivalent to an integral over time for a stream circuit of $\phi = 2\pi/m$; which takes a time $\Delta t = 2\pi/m\,|\Omega - \Omega_p|$. To leading order, we therefore obtain

$$\langle \tau \rangle = \frac{m}{2\pi} \Phi_a F(\zeta) \, {\rm sign}(\Omega - \Omega_p) \quad (33)$$

$$\dot{M} = (\Sigma_{\rm gas} R^2 \Omega) \left[\frac{\Phi_a}{V_c^2} \right] \frac{m\,{\rm sign}(\Omega - \Omega_p)}{(1 + \partial \ln V_c/\partial \ln R)} F(\zeta) \quad (34)$$

$$F(\zeta) \approx {\rm sign}(\zeta) \left(\frac{1}{2} + \mathcal{O}(|\zeta|^{-1}) \right) \quad (35)$$

Note that equations 33-35 in the low ζ limit are comparable to equations 25-28 derived in the $\zeta \gg 1$ limit (to a factor of ~ 2), subject to the additional restriction that equation 30 for the mode amplitude $|a|$ (and thus ζ) must be satisfied in order for there to be shocks in the low ζ limit.

2.3.4 Combined Scaling

For $|\zeta| \approx 1$, we can reasonably interpolate between the two regimes considered in the previous two subsections ($|\zeta| \ll 1$ and $|\zeta| \gg 1$) to find:

$$\dot{M} = (\Sigma_{\rm gas} R^2 \Omega) \left| \frac{\Phi_a}{V_c^2} \right| \frac{m\,S(\omega, \Phi_1)\,F(\zeta)}{1 + \partial \ln V_c/\partial \ln R} \quad (36)$$

$$S(\omega, \Phi_1) = {\rm Re} \left[\frac{m|\Omega - \Omega_p|}{m\Omega - \omega} \right] \cos\left(\arg[\Phi_a \zeta^*] \right) \quad (37)$$

$$F(\zeta) \approx \frac{1}{2} + \frac{|\zeta|}{6 + 2|\zeta|/3} + \frac{|\zeta|^2}{1 + (|\zeta| + 9|\zeta|^2 + |\zeta|^3)/13}. \quad (38)$$

Recall again that in the low ζ limit, these expressions require that equation 30 for the mode amplitude $|a|$ be satisfied.

Note some critical features of these results. The inflow rate $\dot{M}$ is linear in $|a|$ and thus, as advertised, shocks induced in the gas by the non-axisymmetric stellar potential (when present) will dominate the total angular momentum loss relative to the direct transport by the gaseous mode itself. We collect the phase information in the term S in equation 37 ($S < 0$ giving inflow, $S > 0$

outflow). If the modes are purely real and global $S \rightarrow \text{sign}(\Omega - \Omega_p)\,\text{sign}(\Phi_a)\,\text{sign}(\zeta)$. The generalization of S in equation 37 allows for modes with radially varying phases, i.e. $\Phi_a \propto \exp(i \int k\,dr)$, as in §2.3.2. It is easy to show that for a mode in a galaxy with typical structural properties, there is generally inflow induced inside of corotation, and outflow outside (for $\text{sign}(\Phi_a) < 0$, then $\text{sign}(\zeta) > 0$ and $\text{sign}(\Omega - \Omega_p)$ changes from positive to negative from inside to outside corotation); this is consistent with the shock-free, weak-torque results in §2.2. Also for typical structural parameters, the phase changes at the Lindblad resonances, producing outflow inside of the ILR and inflow outside of the OLR.

3 EQUILIBRIUM GAS DENSITIES

In §2, we described the physics that sets the inflow rate of gas due to gravitational torques. Statistically, the *net* inflow rate through a given radius in a galaxy is set by a competition between these torques and star formation. This competition determines the net inflow rate at very small radii (e.g., onto a central BH) and the resulting gas surface density profile.

According to the observed Schmidt-Kennicutt law, the average star formation rate surface density is given by

$$\dot{\Sigma}_* = -\dot{\Sigma}_{\text{gas},*} = \frac{\Sigma_K}{t_K}\left(\frac{\Sigma}{\Sigma_K}\right)^{\eta_K}. \tag{39}$$

where the constants $\eta_K \sim 3/2 - 7/4$, Σ_K and t_K are discussed below. In the limit of a very small supply of gas, the non-axisymmetries in the potential will in general be modest and the inflow of gas will not be able to compete with star formation, which will consume most of the gas. In the opposite limit of a large supply of gas, however, the inflow rate approaches a linear function of Σ_{gas}, while the star formation rate is a nonlinear function. Thus the gas will be consumed rapidly until the inflow and star formation reach an approximate, time-averaged equilibrium. We will consider this in more detail in §5.2 but for now note that inflow and star formation balance each other at a quasi-steady gas density $\langle\Sigma_{\text{gas}}\rangle$ given by

$$\langle\Sigma_{\text{gas}}\rangle \sim \Sigma_K\left(\frac{|\dot{M}|\,t_K}{\pi R^2 \Sigma_K}\right)^{1/\eta_K}. \tag{40}$$

Moreover, because $\dot{M}$ itself depends on Σ_{gas}, equation (40) implies an equilibrium gas surface density *profile*. For the simple dimensional scaling $|\dot{M}| \sim |a|\,\Sigma_{\text{gas}} R^2\,\Omega$, we find

$$\Sigma_{\text{gas}}(R) \sim \Sigma_K\left(\frac{|a|\,t_K}{\pi}\,\Omega(R)\right)^{1/(\eta_K-1)}. \tag{41}$$

Since $\Omega \sim \sqrt{GM_{\text{enc}}(<R)/R^3}$, this implies that the scaling of the *total* surface density sets the equilibrium gas density profile, if we are in the strong torque regime.

We stress that equation 41 and the associated balance between star formation and inflow are only approximately valid, and only then in a time-averaged, order-of-magnitude sense. There are regimes (e.g. the early stages of coalescence in a major merger) where gas can flow in on of order a single dynamical time (Barnes & Hernquist 1991; Mayer et al. 2010), much faster than star formation (observationally supported by observations of e.g. ULIRGs with very high nuclear gas fractions; see Solomon et al. 1997; Bryant & Scoville 1999; Tacconi et al. 1999; Hibbard & Yun 1999). In the opposite extreme, if gas densities are sufficiently low, the star formation and/or inflow timescales can be much larger than the dynamical time. We show below that simulation gas densities typically scatter by an order of magnitude about the result implied by

equation 41. Nonetheless, this is a useful estimate of the gas densities achieved, particularly for analytic purposes. It is also important to note that the equilibrium between star formation and inflow still allows for an arbitrary range in observed gas fractions, as new stars are continuously produced.

To use equation 41 for specific predictions, we require values for Σ_K, t_K, and η_K. From the fit in Kennicutt (1998), $t_K \approx 0.63 \times 10^9\,\text{yr}$ and $\eta_K \approx 1.4 - 1.5$, for $\Sigma_K = 10^8\,M_\odot\,\text{kpc}^{-2}$. However, there is considerable uncertainty. For example, Bouché et al. (2007) suggest $t_K \approx 0.41 \times 10^9\,\text{yr}$ and $\eta_K \approx 1.71 \pm 0.05$, for $\Sigma_K = 10^8\,M_\odot\,\text{kpc}^{-2}$, from observations of high-redshift systems and more extreme starbursting galaxies. Davies et al. (2007) and Hicks et al. (2009) measure star formation in nuclear regions around AGN ($\sim 1 - 100\,\text{pc}$), perhaps closest to the regions of interest here, and find results consistent with the scalings in Bouché et al. (2007). When necessary, we use the latter throughout this paper.

4 COMPARISON TO HYDRODYNAMIC SIMULATIONS
4.1 The Simulations

Thus far, we have made a number of simplifying assumptions in order to derive tractable analytic results. In this section, we compare these analytic models to numerical simulations of gas inflow in galaxies. We do not present any new simulations, but instead draw on results that have been described previously in a number of papers (see e.g. Di Matteo et al. 2005; Robertson et al. 2006b; Cox et al. 2006; Younger et al. 2008; Hopkins & Quataert 2010a). The simulations we compare to include galaxy-galaxy mergers, isolated galaxy disks, and multi-scale "re-simulations" that follow gas from large scales in galaxies to sub-pc scales around a BH. Table 1 summarizes the broad properties of these simulations and provides references to the original papers.

The simulations were performed with the parallel TreeSPH code GADGET-3 (Springel 2005), based on a formulation of smoothed particle hydrodynamics (SPH) that conserves energy and entropy simultaneously even when smoothing lengths evolve adaptively (see e.g., Springel & Hernquist 2002; Hernquist 1993; O'Shea et al. 2005). The detailed numerical methodology is described in Springel (2005), Springel & Hernquist (2003), and Springel et al. (2005). The model galaxies consist of gas disks, stellar disks, stellar bulges, BH particles, and dark matter halos, with Hernquist (1990) halos, exponential disks, and (optional) bulges with mass fractions $B/T \sim 0.1 - 1$. Some of the simulations include sub-resolution prescriptions for BH accretion and feedback, but they are not important for the physics described here – for our results, only the BH mass is critical, and then only on the smallest scales $\lesssim 10\,\text{pc}$ where it dominates the gravitational potential. On all scales, extensive resolution tests have been performed (see references in Table 1), with high-resolution case studies using > 100 times as many particles.

We consider a large number (~ 50) of simulations of isolated galactic disks, marginally to strongly self-gravitating (described in detail in Younger et al. 2008). We also consider a large number of mergers of these idealized disks, in which we vary the merger mass ratio (from 1:20 to 1:1), orbital parameters, and structural properties of the disks (for details, see Hopkins et al. 2009b). These structural variations include changing the disks in mass, bulge-to-disk ratios, disk scale height, gas and stellar disk scale lengths and mass profiles, and scaling the disks following Mo et al. (1998) to correspond to expected disk properties at various redshifts from $z = 0 - 6$. Finally, we also include the "re-simulations" of galactic nuclei described in detail in Hopkins & Quataert (2010a). In these

Table 1. Overview of the Simulations

Simulations	Spatial Scale	Mass Scale [$M_\odot$]	ϵ [pc]	References
Major Mergers	$0.1 - 100\,\mathrm{kpc}$	$10^9 - 10^{12}$	$10 - 100$	Cox et al. (2006), Tables 1-2; Robertson et al. (2006a), Table 1
				Hopkins et al. (2009b), Table 1 & Fig. 16
Minor Mergers	$0.1 - 100\,\mathrm{kpc}$	$10^9 - 10^{12}$	$10 - 100$	Younger et al. (2008), Tables 1-2
Isolated Disks	$0.1 - 10\,\mathrm{kpc}$	$10^9 - 10^{12}$	~ 30	Younger et al. (2008), Tables 3-4; Cox et al. (2008)
Intermediate-scale Re-simulations	$10 - 1000\,\mathrm{pc}$	$10^7 - 10^{10}$	~ 1	Hopkins & Quataert (2010a), Tables 1-2
Nuclear-scale Re-simulations	$0.1 - 100\,\mathrm{pc}$	$10^6 - 10^8$	~ 0.1	Hopkins & Quataert (2010a), Table 3

List of SPH simulations used to test our analytic predictions. From left to right, columns summarize: (**1**) Simulation "class," each of which includes ~ 100 individual simulations, with different structural properties (mass profiles, kinematics) of the initial disks, bulge-to-disk ratios, BH mass, gas fraction, baryonic mass, orbital parameters (in mergers), and efficiency of stellar feedback and star formation. Particle numbers range from $\sim 10^5 - 10^7$. (**2**) Approximate range of spatial scales spanned by each simulation class. (**3**) Corresponding total range in baryonic mass. The particle mass is $\sim 10^{-6}$ times smaller. (**4**) Typical spatial resolution/softening length (often varied with galaxy scale length). (**5**) Reference(s) for the details of each simulation class. Particular tables and figures in the original references that summarize the simulation properties are noted.

calculations, we re-simulate the central $\sim$kpc of a representative sub-set of the mergers and isolated disks at several times in their evolution, at much higher resolution, for a total of ~ 100 simulations. This technique allows us to study, using realistic initial conditions, the formation of secondary instabilities when the gas/stars are self-gravitating ("bars within bars") and the resulting gas inflow to small radii. These simulations are accompanied by a similar number of yet smaller-scale re-simulations that follow the gas inflow deep into the BH potential, with $\sim 0.1\,\mathrm{pc}$ resolution (the resolution studies mentioned above include both varying this spatial resolution and the scales at which the re-simulations are performed; neither of these changes our conclusions).

In both sets of "re-simulations," we consider simulations with zero initial perturbations, with initial perturbations seeded randomly, and with initial perturbations inherited from the "parent" simulation. In Appendix A3 of Hopkins & Quataert (2010a) we show that this does not qualitatively change our results, because the systems of interest are self-gravitating on small scales and so the local structure determines the equilibrium modes, not the initial conditions. Moreover our general analytic model appears to describe shocks due to modes from either initial condition equally well.

The simulations utilized here all include gas cooling and star formation, with gas forming stars at a rate motivated by the observed Kennicutt (1998) relation. Specifically, we use a star formation rate per unit volume $\dot{\rho}_* \propto \rho^{3/2}$ with the normalization chosen so that an isolated Milky-way like galaxy has a total star formation rate of about $1\,M_\odot\,\mathrm{yr}^{-1}$. This formulation produces a projected Schmidt-Kenicutt relation (i.e., $\dot{\Sigma}_*$) similar to that measured by Bouché et al. (2007). We crudely account for feedback from stars with an effective equation of state that mimics turbulence in a multi-phase interstellar medium by adopting a sound speed that is much larger than the true thermal sound speed (e.g., Springel & Hernquist 2003). Hopkins & Quataert (2010a) describe in detail the effects of different subgrid ISM sound speeds on angular momentum transport and inflow rates. They also argue for effective turbulent speeds of $\sim 10 - 50\,\mathrm{km\,s}^{-1}$ for densities $\sim 1 - 10^5\,\mathrm{cm}^{-3}$, respectively. As we have shown in §2, the torques and inflow rate induced in the gas by an asymmetric stellar potential are *independent* of the ISM sound speed in the orbit-crossing limit (e.g., eq. 36). This suggests that the use of a subgrid sound speed in the simulations does not have a significant effect on the physics of gas inflow. However, the critical assumption in *both* the numerical simulations and the analytic calculations presented here is that stellar feedback

maintains a reasonable fraction of the ISM in a diffuse phase, rather than being bound in star clusters. In the simulations, gravitational instability can form some compact, dense clumps (analogous to molecular clouds or star clusters), but the enhanced subgrid sound speed ensures that most of the ISM does not suffer this fate. This is intended to mimic physics not included in our calculations, such as the observed low efficiency of star formation in dense molecular gas (e.g., Krumholz & Tan 2007) and the disruption of molecular clouds by radiation pressure, HII regions, etc. (e.g., Murray et al. 2010) – the latter returns most of the gas from molecular clouds back into the diffuse ISM.

There are of course considerable uncertainties in any model of the ISM and star formation in galaxies, especially on small scales near a BH; the adoption of a simple Kennicutt (1998) relation is merely a convenient simplification. For this reason, Hopkins & Quataert (2010a) consider a broad range of such prescriptions (see their § 2 and Appendices). We briefly summarize some of these variations, but interested readers should see that paper for details. The slope of the star formation law was varied from $\dot{\rho}_* \propto \rho^{1-2}$, amounting to changes in the star formation efficiency on small scales by factors of $\gtrsim 100$. This spans the entire observed parameter space of star formation efficiencies on large galactic scales (see e.g. Bigiel et al. 2008), in dense molecular clouds (Krumholz & Tan 2007), and the central regions of AGN (Hicks et al. 2009); our "typical" cases are consistent with the latter observations. The primary difference the star formation efficiency makes is that it changes the gas-to-stars ratios at a given time; but since this quantity explicitly appears in our model, the effect is to move systems along our predicted scalings, not with respect to them. In our analytic estimates that follow, when the absolute value of star formation efficiency is important, we parameterize it as a free function, allowing for different detailed models of star formation.

The simulations in Table 1 also include large systematic studies of the prescription for the sub-grid gas physics in the ISM. Specifically, the effective equation of state of the gas was widely varied. The simulations include cases where it is quite stiff on small scales (near-adiabatic, similar to what is adopted in the studies of Mayer et al. 2007; Dotti et al. 2009), through to cases where the gas is allowed to cool to a cold isothermal floor (with $c_s \ll V_c$), and thus forms a clumpy, inhomogeneous medium (on galactic scales, similar to what is assumed in Bournaud et al. 2007; Teyssier et al. 2010). Intermediate cases are also sampled, allowing for e.g. an equation of state similar to that presented in Klessen et al. (2007), motivated by the models of Spaans & Silk (2005) for dense star-forming gas. In a large number of our nuclear-scale simulations, the gas disper-
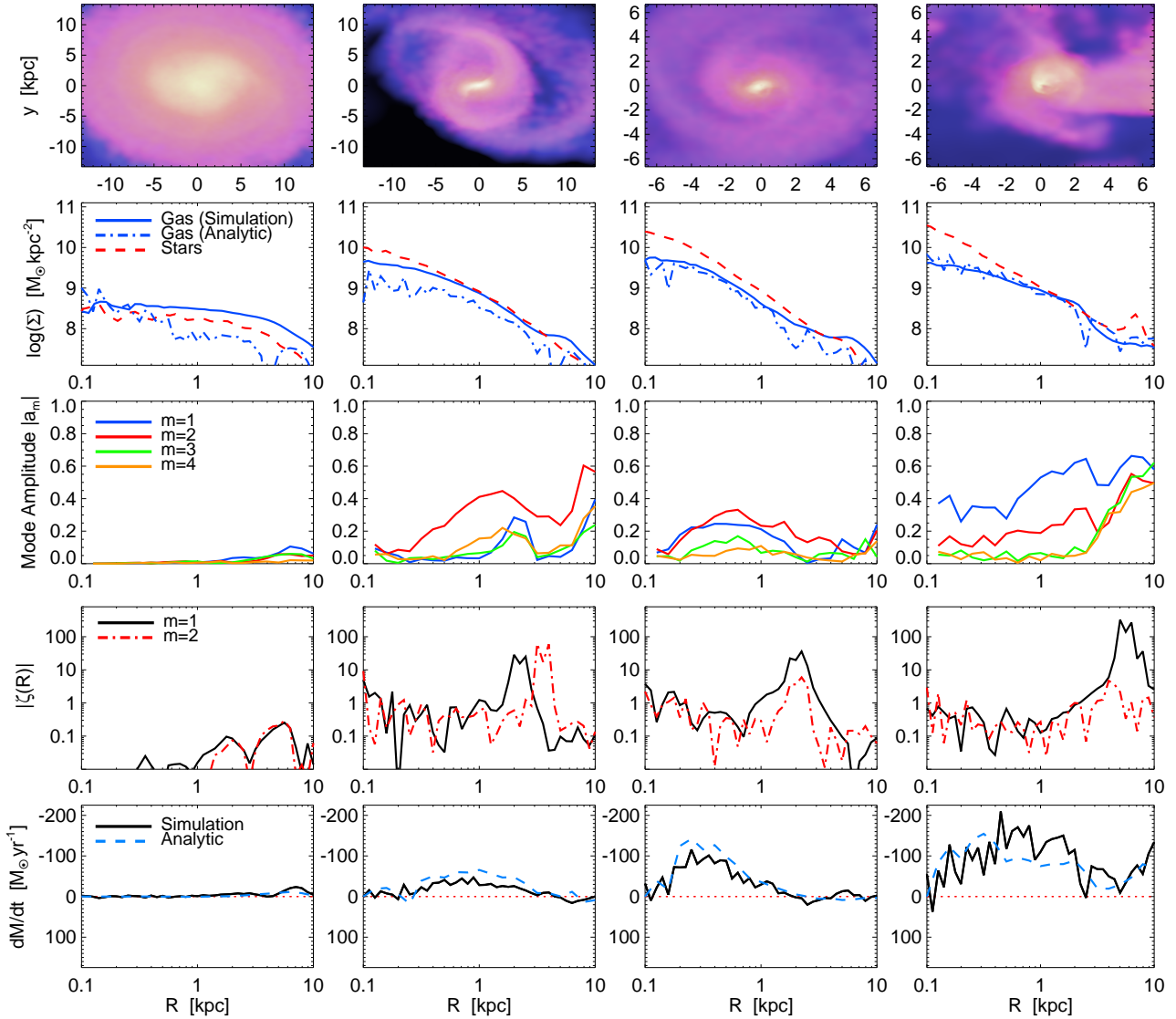
Figure 1. Inflow properties in a typical gas-rich ($f_{\rm gas} \approx 0.4$) merger of two equal-mass $\sim L_*$ ($M_* = 10^{11} M_\odot$) galaxies (run b3ex in Hopkins & Quataert 2010a), at different times: before any interaction, when the system is unperturbed (*left*); just after an early first passage (*center left*), which induces a bar response in the primary; somewhat later, just before second passage, as the induced bar feature has relaxed (*center right*); just after final coalescence (*right*). *Top:* Images of the gas density at each time. Intensity reflects the gas surface density. Color reflects the specific star formation rate (from blue being low, with gas consumption timescale $\gg$Gyr to yellow being high, with consumption timescale $\sim 10^7 - 10^8$ yr). Horizontal axis is cartesian (x, in kpc). *Second from Top:* Gas and stellar surface density profiles as a function of cylindrical radius R. We also show the analytic gas density profile from eq. 41, which assumes that gas inflow approximately balances star formation at each radius (see § 3). *Middle:* Amplitude of non-axisymmetric modes versus radius. *Second from Bottom:* Orbit-crossing measure $|\zeta(R)| \equiv |\partial R_1/\partial R|$ (eq 20); $|\zeta| > 1$ implies strict orbit crossing, but shocks are still present at many radii with $|\zeta| \lesssim 1$. *Bottom:* Gas inflow (negative) or outflow (positive) rate as a function of radius. Dotted red line shows $\dot{M}_{\rm gas} = 0$. We compare the simulation results to our analytic prediction based on strong shocks induced in the gas by the non-axisymmetric stellar potential (§ 2.3.4). The predictions from the no-shock, weak-torque limit (eq 17) or the direct transport by the mode in the gas (eq 18) are smaller by a factor > 10, and are would be indistinguishable from the $\dot{M} = 0$ line.

sion is dominated by *resolved* turbulent gas motions, which maintain an effective c_s even as the sub-grid c_s becomes very small (the origin of this turbulence and its importance for e.g. AGN torii will be discussed in future work). As such, we can explicitly check that our formulation for gravitational torques still applies even in turbulent inhomogeneous gas distributions.

Nevertheless, we consider our sub-grid treatment of the ISM to be the biggest uncertainty in the numerical models. Because we cannot resolve the extremely large densities of actual star-forming cores in these regions, star formation is almost certainly more diffuse/uniform throughout the disk, which may in turn have conse-

quences for the ability of stars and gas to interact efficiently. And the ISM is, by definition, more smooth than the implicit sub-grid multiphase structure. An important goal for future work will be the investigation of more realistic ISM and star formation models on these scales. These are unlikely to make a large difference in the limit of stellar-dominated disks (our focus here), but may in the very gas-rich regime open up new channels for angular momentum exchange.

In §4.2 we provide an overview of the simulation results and a comparison to the analytic results derived in §2 & 3 for example simulations covering three different spatial scales: galactic ($\sim$
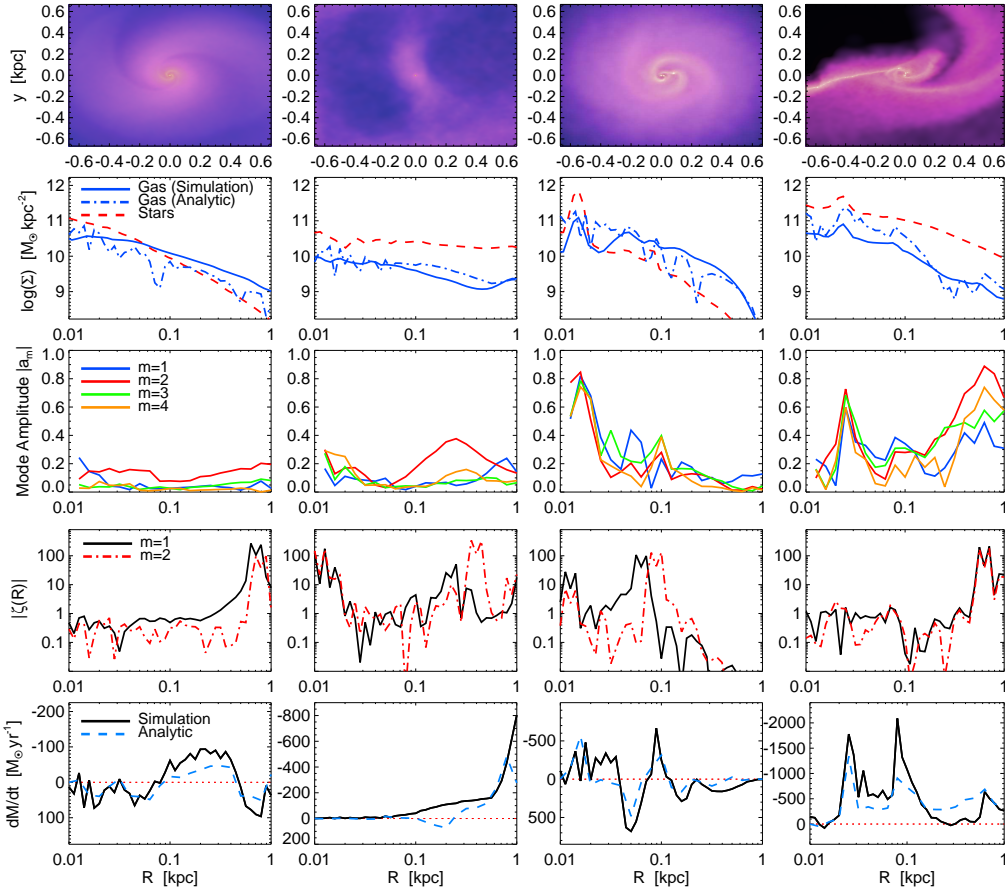
Figure 2. Analogous results to Figure 1 for re-simulations of the central ~kpc in galaxy mergers, at representative times in the large-scale inflow history (from left to right, these are runs If9b5res, If1b1late, Ilowresq, Inf28b4 in Hopkins & Quataert 2010a). These simulations resolve gas inflows to within ~ 10 pc.

kpc), intermediate-scale (~ 0.01 − 1 kpc), and nuclear (~ 0.1 − 10 pc). We refer the reader to Hopkins & Quataert (2010a) for more detailed information about the properties of the gas and stars in the simulations (e.g., Q[r], h/r, etc). Although the specific simulations discussed in §4.2 are quite instructive, it can also be misleading to focus on the results of an individual simulation. The reason is that there is large variation in time and potentially large scatter introduced by modest differences in galaxy properties (due to, e.g., large-scale fragmentation of a spiral arm biasing the results of a given simulation). For this reason we believe that it is important to consider the statistical properties of a large number of simulations. In § 4.3 we thus present a more comprehensive statistical comparison between the simulations and our analytic results.

4.2 Overview of Simulation Results

Figure 1 provides an overview of the dynamics during a major merger of equal-mass ~ L_* galaxies with $f_{gas} \approx 0.4$. The first row shows, from left to right, images of one of the two galaxies before first passage, after a near-final passage (when the impact parameter of the interaction is $\lesssim$ the size of the galaxy), and just before the final coalescence of the two systems; the final column shows an image of the remnant system just after coalescence. The second row in Figure 1 shows gaseous and stellar mass surface density profiles at each time, while the third shows the amplitude of the stellar surface density perturbations for different azimuthal mode number m

as a function of radius ($R = 0$ is defined as the location of the BH);

$$|a_m(R,t)| = \frac{|\int_0^{2\pi} \Sigma(R, \phi) \exp(\imath m \phi) \, d\phi|}{\int_0^{2\pi} \Sigma(R, \phi) \, d\phi} \quad (42)$$

The phase angle ϕ_0 of the mode maximum is determined using a similar expression, as are the potential perturbations, and other perturbed quantities of interest. The second to last row in Figure 1 shows estimates of the orbit crossing parameter ζ for the $m = 1$ and $m = 2$ modes (the calculation of which is described below). Finally, the bottom row in Figure 1 shows the instantaneous gas inflow rate as a function of radius R, determined directly as $\dot{M} = \Delta R^{-1} \int dM_{gas} v_R$ given the radial velocity v_R of each gas particle of mass dM_{gas} in a small radial annulus ΔR. As before, we define $\dot{M} < 0$ to be accretion/inflow, whereas $\dot{M} > 0$ corresponds to outflow.

In addition to the numerical results, Figure 1 also shows several of the predictions of our analytic model. First, given the measured inflow rate in the simulations $\dot{M}$, we use the balance between inflow and star formation implied by equation 40 to predict the gas surface density Σ_{gas} (2nd row). This prediction is within a factor of ~ 3 of the measured gas density in the simulations, particularly at later times (right-most two columns) when our assumed balance between inflow and star formation is a better approximation. In addition, given the measured $|a|$ at each radius, and the local gas properties (Σ_{gas}, Ω, and their derivatives), Figure 1 also shows our prediction of the inflow rate $\dot{M}$ using equation 36 (bottom row). Evaluating the sign of $\dot{M}$ in equation 36 (i.e., inflow or outflow), is non-trivial given its dependence on Ω_p and ζ; one option would
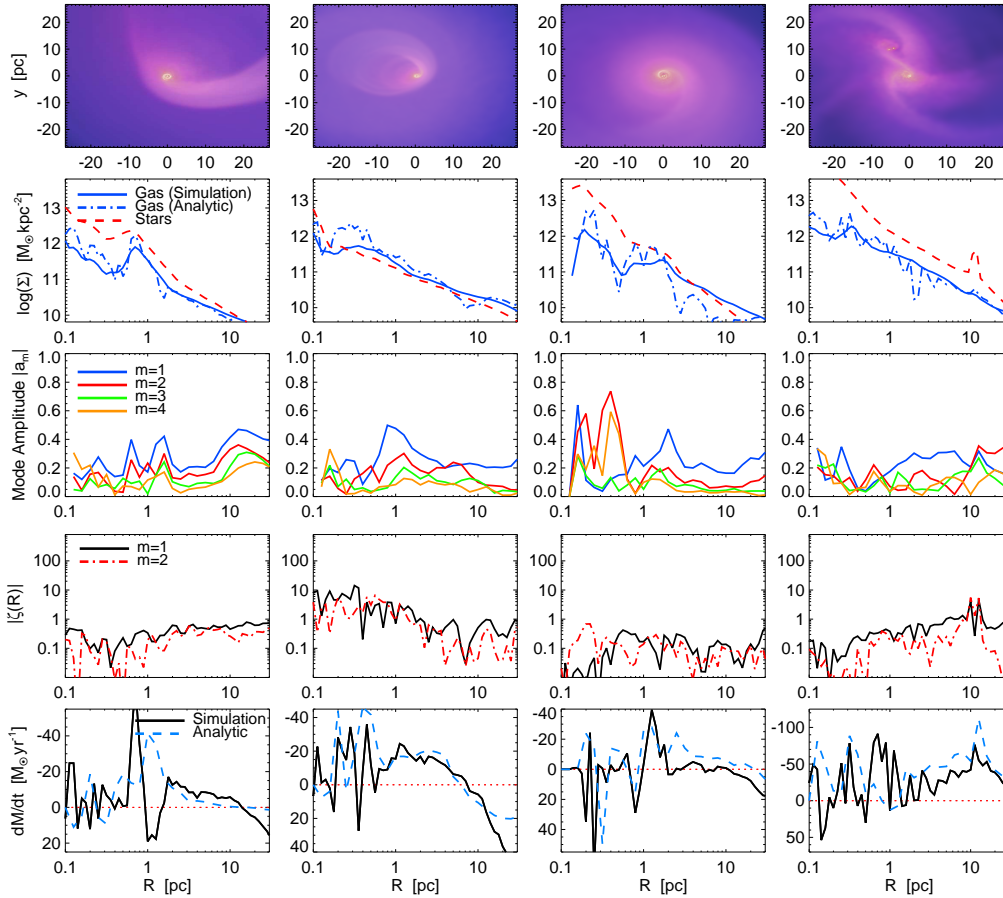
Figure 3. Analogous results to Figure 1 for re-simulations of the central $\sim 10 - 50\,\mathrm{pc}$ in the intermediate-scale simulations shown in Figure 2, at times near the peak of accretion and gas inflow (from left to right, these are runs Nf8h1c1qs, Nf8h2b4, Nf8h1c0hol, Nf8h1c1dens in Hopkins & Quataert 2010a). These simulations have a resolution $\sim 0.1\,\mathrm{pc}$ and extend deep into the potential of the central BH. The (non-wound) $m = 1$ eccentric disk and/or single-armed spiral mode is ubiquitous in the quasi-Keplerian potential.

be to fit a global model to the mode structure as a function of time and radius, but since ζ depends on second derivatives this is very noisy. Instead, we found it easier and more numerically robust to use the epicyclic approximation to estimate ζ and the sign of $\dot{M}$. With this approximation, the mean local radial velocity (zero in the unperturbed state) obeys

$$\frac{\partial v_{1,R}}{\partial R} \sim i\, m\, (\Omega - \Omega_p)\, \zeta, \qquad (43)$$

i.e. it is phase-shifted by $\pi/2$ and multiplied by $(\Omega - \Omega_p)$ relative to ζ; this implies that $\mathrm{sign}(\Omega - \Omega_p) \cos{(\arg[\Phi_1 \zeta^*])} = -\sin{(\arg[\Phi_1] - \arg[\partial v_1 / \partial R])}$. We use the true radial velocity of the stars in the simulations v_R as our estimate of the perturbed velocity $v_{1,R}$ and determine its amplitude and phase as in equation 42. Together with the measured Φ_1 in the simulations we can thus determine the sign of $\dot{M}$ induced by a given m. For the results in Figure 1, we sum the contributions to $\dot{M}$ from $m = 1$ to $m = 5$, but find that this is almost always dominated by the largest mode or two (this will be an assumption of ours later). Equation 43 also implies that $|\zeta| \sim |\partial v_{1,R} / \partial R| / [m\, |\Omega - \Omega_p|]$; we estimate Ω_p from a global fit to the mode amplitude and axis ϕ_0 as a function of time, assuming Ω_p is constant in time and radius for a given simulation. The results for $|\zeta|$ are shown in the 2nd to last row in Figure 1 for the major merger simulations.

Figure 1 shows that the predicted inflow rate as a function of radius is well reproduced by our analytic model at all times,

providing support for the simplifying assumptions we made in the analytic derivation. By contrast, the predictions of the no-shock, weak-torque limit (eq 17) or the direct transport by the mode in the gas (eq 18) are smaller by a factor > 10. Although the results in Figure 1 are for a major merger, we find similar consistency between the simulation results and our analytic predictions for both minor mergers and isolated, bar-unstable galactic disks. Figure 1 also shows that the orbit crossing parameter ζ is ~ 1 at a range of radii, with $|\zeta| \sim 10 - 100 \gg 1$ near the co-rotation resonance, implying strong orbit crossings and shocks. As described in the text above equation 43, our estimate of ζ is itself based on the epicyclic approximation; it is thus very approximate in mergers given that the galaxies are strongly disturbed and quantities such as Ω_p are not that well-defined. Nonetheless, the results in Figure 1 are consistent with the presence of strong shocks and dissipation in the simulations, which drive the large gas inflow rates.

Figures 2 and 3 show analogous results to Figure 1, but for simulations of the central $\sim$kpc (Fig. 2) and ~ 10 pc (Fig. 3) of gas-rich mergers (from Hopkins & Quataert 2010a). The simulations span different morphologies and degrees of clumpiness in the ISM gas, and include both cases that are strongly stellar-dominated and that have more comparable gas and stellar densities. In each case, the agreement between the analytically predicted Σ_{gas} and $\dot{M}$ and the numerical results is again quite good. As noted above, we do not necessarily expect our predictions to be applicable in
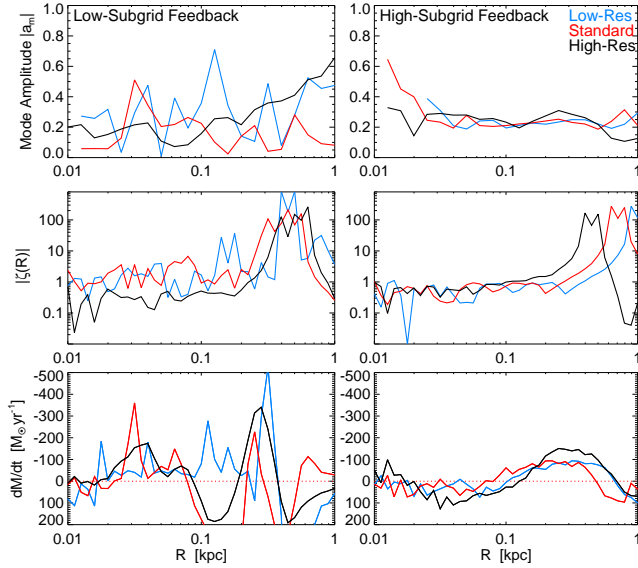
Figure 4. Effects of simulation resolution on the modes we compare to our analytic model. We plot the $m = 2$ mode amplitude (*top*), $|\zeta(R)|$ for the $m = 2$ mode (*middle*), and inflow/outflow rate (*bottom*), for intermediate-scale simulations as in Figure 2. For each, we show three otherwise identical simulations: one with low resolution ($N_{particles} = 5 \times 10^5$; comparable to the resolution in the central regions of our galaxy-scale runs before "zoom-in"), one with our standard resolution ($N = 5 \times 10^6$; that of the other zoom-in runs in Figure 2), and one with very high resolution ($N = 5 \times 10^7$; high enough to obviate the need for an additional zoom-in at $< 10\,$pc scales). We compare a model with very weak feedback (*left*; $q_{eos} = 0.02$ giving an effective ISM sound speed $c_{eff} \sim 10\,$km s^{-1}) to one with strong feedback (*right*; $q_{eos} = 0.25$, or $c_{eff} \sim 50\,$km s^{-1}). The former shows more sensitivity to resolution because of sub-structure that forms in the modes; the latter gives a more smooth ISM so converges more quickly. At the order-of-magnitude level, however, the mode amplitudes and inflow rates are similar.

gas-dominated systems; however, at least up to gas fractions ~ 0.5 seen here, the inflow rates remain well-described by our formulation. And the Figures include both simulations with initial conditions set to have "smooth" structure (no initial modes) and those with a wide spectrum of initial modes inherited from the structure of their "parent" simulations. More than Figure 1, Figures 2 and 3 demonstrate that our analytic model even reproduces the sign of $\dot{M}$ (inflow vs. outflow) reasonably well. Note also that the $m = 1$ mode becomes increasingly important relative to the $m = 2$ mode at small radii (Fig. 3). This is not only clear visually in the images, but also quantitatively in the mode amplitudes.[1] At the smaller radii shown in Figures 2 and 3, the modes are more often in the marginal orbit-crossing limit, with $|\zeta| \sim 0.1-$a few, rather than $\zeta \gg 1$; this highlights the importance of our results in § 2.3.3, which show that shocks and rapid inflow can occur even where $|\zeta| \lesssim 1$ (although not for $\zeta \ll 1$).

Figure 4 briefly illustrates how the simulation properties depend on the resolution and uncertain microphysics of the numerical models. We show the mode amplitudes and $|\zeta(R)|$ (for simplicity restricting to just $m = 2$), and inflow rates as a function

of radius, for simulations of the central $\sim$kpc (the range shown in Fig. 2). We consider otherwise identical initial conditions but with different resolution. We also consider two different extremes in the parameterization of the ISM gas equation of state, encapsulated in the parameter q_{eos} (see Hopkins & Quataert 2010a for details). A low value of q_{eos} corresponds to an ISM with efficient cooling and without much turbulent or thermal pressure support (here $q_{eos} = 0.02$ corresponds to an effective sub-grid sound speed of just $\sim 10\,$km s^{-1}), and a high value corresponds to an ISM with strong pressure support (here $q_{eos} = 0.25$ corresponds to effective sound-speeds $\sim 50\,$km s^{-1}). For each, we compare simulations at three resolution levels: a low-resolution run, for which the mass and force resolution inside of $\sim 1\,$kpc are comparable to what can be achieved directly (without any "zoom-in" or "re-simulation") in the parent, galaxy-scale simulations; a standard-resolution run (the characteristic resolution of our re-simulations on this scale); and a very high-resolution run (the "ultra-high" resolution runs in Hopkins & Quataert 2010a) with $\sim 0.3\,$pc force resolution, small enough to resolve the same scales as our nuclear-scale simulations without the need for an additional re-simulation iteration. Even at low resolution, the major qualitative features are in place and the inflow rates are order-of-magnitude accurate. But in the case with weak assumed feedback, the ISM is subject to much greater resolved fragmentation and clumpy star formation (see Appendix B in Hopkins & Quataert 2010a for more discussion and examples). This leads to inflow rates and mode structures that are more highly variable in position and time, and the exact details (depending on how clumps collapse locally and turn into stars) are more sensitive to resolution. The case with stronger implicit feedback yields an ISM with less explicit substructure, which in turn allows rapid convergence with resolution. But in both low and high-feedback cases, the average behavior can still be reasonably approximated with the analytic prescriptions (accounting for the implicit, sub-grid ISM structure assumed in the numerical models). Because of the uncertain details of star formation and ISM physics in the simulations, we do not think any of these runs should be taken literally as an exact description of what really happens in AGN; however, they provide a means to test our analytic prescriptions in fully non-linear, chaotic systems.

Note that on small scales, we have chosen times after the two BHs in a galaxy merger are assumed to merge. It is of course not clear if the final merger proceeds quickly or slowly on sub-pc scales. We do so for simplicity and generality (since isolated disks will not have a second black hole). But during the actual binary phase of a near equal-mass BH-BH merger, the mode structure and gravitational torques on small scales will be highly non-linear and feature a more complex geometry than what we show here. We can, however, still attempt to consider this in the context of our analytic framework, but with the secondary BH as the "collisionless" driver of non-circular gas motions and shocks, as opposed to modes in the stellar disk. For more discussion on the inspiral phase in these simulations and its implications for the presence of the $m = 1$ modes discussed above, we refer to Hopkins & Quataert (2010a).

4.3 Inflow Rates

We now present a statistical comparison between the inflow rates in our numerical simulations and the predictions of inflow produced by stellar torques (§2). In doing so, we shall show a number of Figures that contain the results of many of our simulations. In these Figures, the critical point to focus on is less the results of any given simulation (which can be difficult to identify in the Figure), but rather the ensemble properties of all of the simulations (e.g., me-
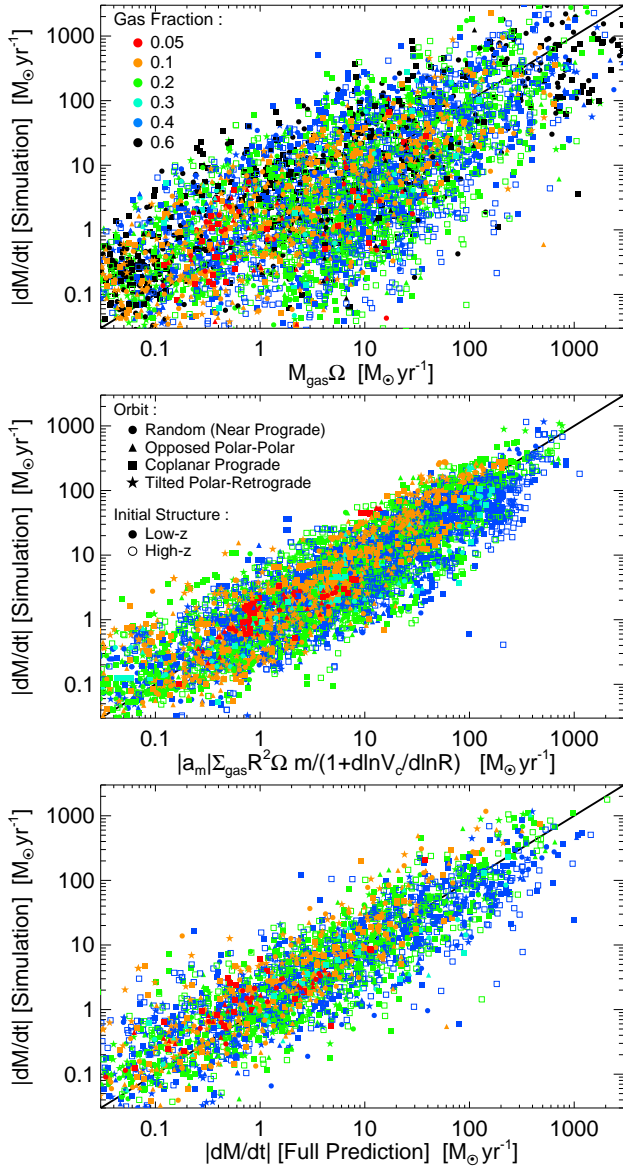
---

[1] The reasons for this are discussed at length in Hopkins & Quataert (2010a) and Hopkins & Quataert (2010b), but in general owe to the fact that $\kappa \approx \Omega$ in a quasi-Keplerian potential; "slow" $m = 1$ modes are therefore generic (though not unique) to small scales near a BH (see Tremaine 2001 for stellar-dominated systems, or Ostriker et al. 1992 for gaseous disks).

Figure 5. Instantaneous gas inflow/outflow rates $|\dot{M}|$ in major galaxy merger simulations versus several predictors of the inflow rate. The simulations span a range of masses, gas fractions, sizes, and orbital parameters. Some of these parameters are explicitly labeled (e.g., color denotes gas fraction, open (filled) points denote initial disks with structural parameters similar to local $z \sim 0$ (high-redshift) galaxies). We calculate $|\dot{M}|$ through a radius $\approx 5\epsilon$ (where ϵ is the simulation resolution limit; results are similar for radii $R \sim 0.1 - 1\,$kpc), at random times uniformly sampling between first passage and 0.5 Gyr after peak activity. *Top:* Simulation results compared to the dimensional scaling $M_{\rm gas}(< R)\,\Omega(R)$. The scatter is very large, $\sim 1.5 - 2\,$dex. *Middle:* Simulation results compared to the simple estimate of inflow induced by strong orbit crossing in the gas (eq. 19). Predicted and simulated inflow rates agree well, with a scatter of $\sim 0.5\,$dex. *Bottom:* Simulation results compared to the full scaling derived for the strong-torque regime (eq. 26). The scatter is reduced to $< 0.3\,$dex.

dian, scatter, or systematic trend relative to analytic predictions). In what follows, we take $F(\zeta) = 1$ for all but the $m = 1$ modes at $\lesssim 10$ pc, for which we take $F(\zeta) = 1/2$; these choices are motivated by the typical values of ζ shown in Figures 1-3.

Figure 5 compares the instantaneous inflow rate in major galaxy merger simulations to several analytic predictors of the in-

flow rate: the simplest possible dimensional expectation, $M_{\rm gas}(< R)\,\Omega(R)$ (top panel); the order of magnitude prediction of inflow produced by stellar torques (middle panel; equation 19); and the more detailed prediction of our analytic model (bottom panel; equation 36 from §2.3.4). In all three panels, we measure the inflow rate through a radius of 5ϵ (where ϵ is the gravitational softening length of the simulation, roughly chosen so that $5\epsilon \sim R_{\rm eff}/10$), although we find similar results at other radii. We calculate $\dot{M}$ for each simulation at $\sim 10^8$ yr intervals, from just before first passage until $\approx 0.5\,$Gyr after the peak of activity (when the galaxies fully coalesce), at which point significant inflow has ceased. The instantaneous inflow rate is averaged over each time interval between points to average over both the chaotic motions of the gas and noise introduced by the finite simulation resolution. The mode amplitude and other properties of the gas needed in equations 19 and 36 are calculated from the simulations themselves (at the same radius as the measured inflow rate) as in §4.2.

Figure 5 (top panel) shows that although there is a correlation between the true inflow rate and $M_{\rm gas}(< R)\,\Omega(R)$, the scatter is very large $\sim 1 - 2\,$dex. Note that the large vertical scatter here (particularly that above the prediction) does not imply material falling in on much less than the dynamical time; instead, it stems from a combination of time variability and scatter in the x-axis. The bottom panels of Figure 5 show that the analytic predictions in the strong torque regime, equations 19 and 36, work remarkably well, predicting the instantaneous inflow rates with a scatter of less than $\sim 2 - 3$. Moreover, there are no systematic trends with simulation properties such as the gas fraction, galaxy size, orbital parameters, or ISM equation of state model.

Figure 6 compares the inflow rate to equation 19 (the order of magnitude strong torque $\dot{M}$) for our other sets of simulations: minor mergers, isolated bar-unstable disks, intermediate (sub-kpc) scale re-simulations of galactic nuclei, and nuclear-scale (< 10 pc) re-simulations of disks around BHs. In each case we again measure the inflow rate at $\sim 5\epsilon$ in the simulations and smooth the results in time. The strong torque analytic model works well in all cases, despite a very large dynamic range in mass, spatial scale, and inflow rate, and despite the fact that different, independent non-axisymmetric modes control angular momentum loss at different radii. In addition, some of the simulations (particularly the mergers) are clearly in the strong orbit-crossing regime ($\zeta \gg 1$), while others (particularly the < 10 pc circum-BH disks) are in the marginal orbit-crossing regime ($\zeta \lesssim 1$) in which there are fewer (if any) formal orbit crossings. The same inflow prediction nonetheless works well in both limits. Gas fractions are explicitly labeled by color; although we expect our assumptions to break down somewhere around $f_{\rm gas} \sim 1$, it is immediately clear from Figure 5 that up to $f_{\rm gas} \sim 0.5$, there is no systematic deviation from our scaling.[2] Other parameters varied from simulation to simulation are generally not shown, but we have checked to see that there is no clear systematic difference between our predictions and the results of a given subset of the simulations. For example, the effective equation of state of the gas is varied widely at each scale as discussed in § 4.1, but there is no statistically significant offset of the inflow rates with respect to the predictions in each case (with the important caveat that we have no simulations with a true non-sub grid

---

[2] We emphasize, however, that at gas fractions this high, the sub-grid ISM model of the simulations becomes particularly suspect (as small-scale clumping/fragmentation is not explicitly treated). It is safer to say that our results are robust to variations in $f_{\rm gas}$ for $f_{\rm gas} \ll 1$.
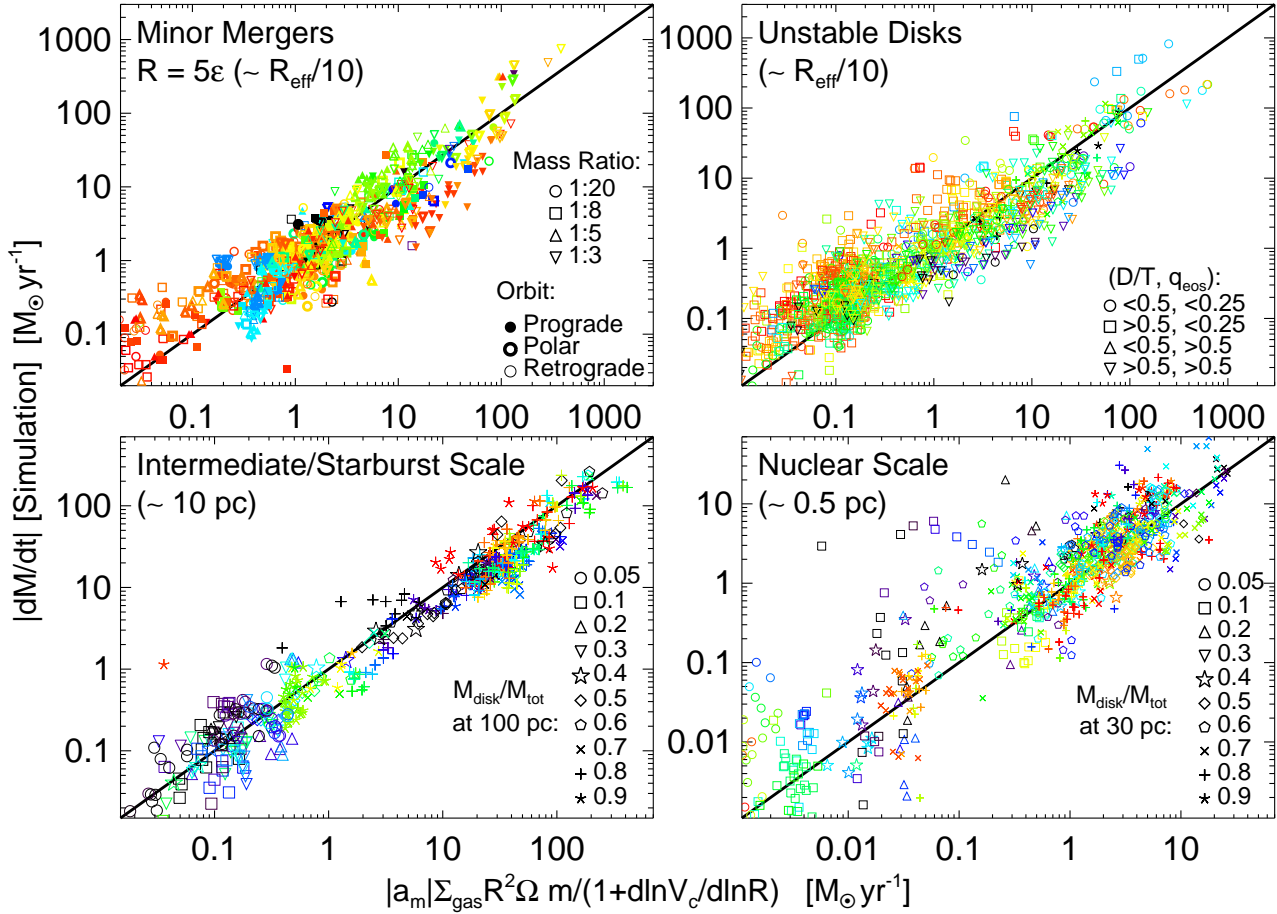
Figure 6. Simulated vs. predicted inflow rates, as in Figure 5, for minor mergers, isolated galactic disks, and re-simulations of galactic nuclei; the inflow rates are measured at a radius of 5ϵ (as Figure 5) in each case, with the values of this radius in parentheses in each panel. Colors denote the gas fraction, as in Fig. 5, while symbol types denote some of the key parameters varied in the ensemble of simulations. In all panels, the predicted inflow rate (x-axis) is the simple strong orbit crossing estimate in equation 19. *Top Left:* Galaxy-galaxy minor mergers. *Top Right:* Isolated, bar-unstable galactic disks. The parameters varied in the simulations include the initial disk-to-total mass ratio D/T at the effective radius and the stellar feedback prescription (q_{eos}, where < 0.25 denotes modest feedback and > 0.5 denotes strong feedback; see Hopkins & Quataert 2010a). *Bottom Left:* Intermediate-scale re-simulations of both mergers and isolated unstable systems (e.g., Fig. 2). The key parameter regulating the formation of secondary bars is the disk-to-total mass ratio. *Bottom Right:* Nuclear-scale re-simulations of the intermediate-scale systems (e.g., Fig. 3), where the lopsided $m = 1$ disk forms and regulates the gas inflow.

model). Likewise, both in isolated disks and nuclear-scale simulations, the model of star formation has been varied in a large subset of simulations; the resulting points are indistinguishable. Making star formation less homogeneous by raising the star formation density threshold does not appear to change our conclusions, but this is again with the caveat that the simulations do no explicitly model individual GMCs. Changes to the ISM equation of state or star formation law do affect various global properties of the results – the mass distribution, gas fractions, etc – but these are explicitly included in our formulation of $\dot{M}$; the important point is that they introduce no significant deviations from our predicted scalings. Parameters such as the initial simulation redshift enter only indirectly, changing disk structural parameters, and Figure 5 shows explicitly that they make little difference to our conclusions.

One apparent outlier in Figure 6 is the "spray" of several (blue/black) points at large true $\dot{M}$, but small predicted $\dot{M}$, in the nuclear-scale simulations (bottom right). This is a system in which the gas disk undergoes a large-scale fragmentation and a gas-rich clump happens to fall on a nearly radial orbit towards the BH. This is a special case. Because of the high density of points in Figure 6, this small number of outliers makes it appear as if the scatter in

this panel in much larger – in fact, the correlation is just as significant as in each of the other panels (linear correlation coefficient 0.87, or significance $> 10\,\sigma$), and the intrinsic scatter is nearly identical (lognormal 0.37 dex, as compared to 0.33 dex for the points plotted in other panels). Recall as well that in many of the simulations shown here, the gas motions include a significant contribution from resolved 'turbulent' motions and the gas mass distribution can be highly inhomogeneous/clumpy; we nonetheless still find good agreement with the predictions of the strong torque analytic model.

Figure 7 shows the same simulation results as Figure 6, but compares them to the predicted inflow in the weak-torque regime, equations 13 & 17. These scalings do fine at high $\dot{M}$, but fail to reproduce the numerical results at lower $\dot{M}$ and are clearly systematically discrepant. If we use the results of Lynden-Bell & Kalnajs (1972) for direct angular momentum transport by the non-axisymmetric modes in the gas (also second-order in $|a|$; see eq. 18), we obtain a similar, but more severe, discrepancy. This is because the inflow rate induced by the direct transport by non-axisymmetric modes (eq. 18) is smaller than the weak-torque inflow produced by the stars on the gas (eqs 13 & 17) by at least a factor of $\sim \Sigma_{gas}/\Sigma_{tot}$ (see §2.2). Overall, Figure 7 highlights that
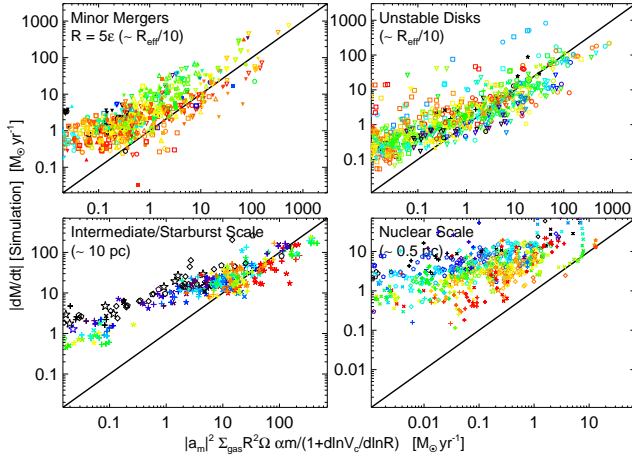
Figure 7. Simulated vs. predicted inflow rates, as in Figure 6, but for the linear weak-torque, non-orbit crossing predictions (eq. 17 with $\alpha(\Lambda) \sim 1$). This prediction works reasonably well at the highest $\dot{M}$, because the second-order formula in $|a|$ approaches that for strong torques (in Fig. 6) when $|a| \to 1$. However, over the majority of the dynamic range in $\dot{M}$ the weak-torque prediction significantly under-predicts the true inflow rate. Note that the axes here are the same as in Figure 6. As a result many of the low simulated $\dot{M}$ points (y-axis) do not show up in this Figure because they have very small predicted $\dot{M}$ (x-axis) and are thus off the panel to the left.

for the entire mass and radial range we have simulated, the dominant torque is well-modeled as originating from shocks induced in the gas by the stars.

5 BRIDGING THE GAP: FROM GALAXY TO BH

The literature contains a number of estimates of the accretion rate onto a central massive BH and/or through a given radial annulus, none of which properly account for the physics we have elucidated here. Here we briefly review several of these estimates (§5.1) and then present a new analytic model for the inflow rate onto a central BH given the conditions at larger radii in its host galaxy (§5.2).

5.1 Previous Accretion Models

In semi-analytic and numerical models that lack the resolution to study gas transport from $\sim 0.1 - 1$ kpc to $\lesssim 0.1$ pc, simple subgrid prescriptions are necessarily used to estimate the accretion rate onto a central BH. One common estimate of the BH accretion rate is the Bondi accretion rate for spherical accretion (e.g., Di Matteo et al. 2005):

$$\dot{M}_{\rm Bondi} = 4\pi \, \alpha_{\rm B} \, \frac{G^2 \, M_{\rm BH}^2 \, \rho_{\rm gas}}{c_s^3} \qquad (44)$$

where c_s and $\rho_{\rm gas}$ are the ambient sound speed and gas density, respectively, and $\alpha_{\rm B}$ is a constant that depends on the mass profile and gas equation of state; $\alpha_{\rm B}$ is sometimes taken to be as large as ~ 100 (Springel et al. 2005). Although perhaps appropriate for pressure-supported gas (e.g., in clusters of galaxies), the Bondi rate is not well-motivated for the cold, nearly Keplerian, gas that dominates the mass in disk galaxies. Indeed, we shall see that equation 44 is not a good approximation to our numerical results.

For rotationally supported gas, the inflow rate is determined by the rate of angular momentum transport. If angular momentum transport in galaxies were primarily produced by local stresses in a thin disk, the effective kinematic viscosity would be $\sim \alpha_V c_s h$,
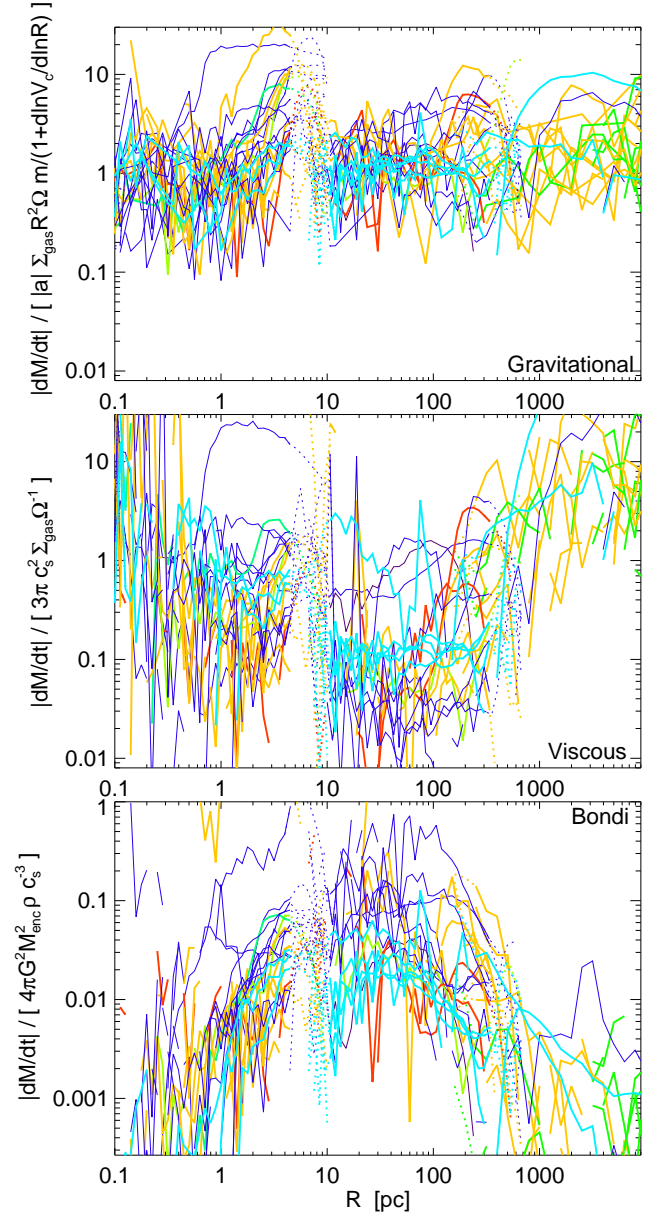

Figure 8. Comparison of simulated and predicted gas inflow rates as a function of radius. The simulations are shown during the strong inflow phases, as described in Figure 5. *Top:* The inflow rate in the simulations at each radius (dM/dt) compared to the simple estimate for strong gravitational torques (eq. 19). Each solid line denotes a different simulation, with dotted lines showing the radii near the boundaries of our re-simulations, where the exact profile shape is suspect; colors correspond to initial gas fractions as in Figure 5. *Middle:* Simulated inflow rate versus the scaling expected for accretion via local viscous stresses (eq. 45). *Bottom:* Simulated inflow rate versus a Bondi-Hoyle-like estimate (eq. 44 with $M_{\rm BH} \to M_{\rm enc}(< R)$). The Bondi-Hoyle and viscous accretion rate formulations do not accurately capture the dominant physics in the numerical simulations, while the strong gravitational torque estimate does quite well.

yielding an inflow rate

$$\dot{M}_{\rm viscous} = 3\pi \, \alpha_{\rm V} \, c_s \, \Sigma_{\rm gas} \, h \qquad (45)$$

where $h \sim c_s/\Omega$ is the disk's vertical thickness and $\alpha_{\rm V}$ is the dimensionless viscosity parameter (assumed spatially constant).

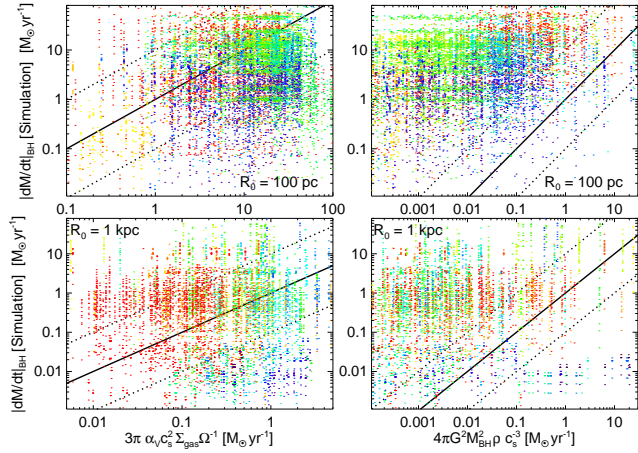Both equations 44 and 45 depend quite differently on the am-

Figure 9. Comparison of the BH inflow rate in our simulations with the viscous (*left*; $\alpha_V = 0.1$) and Bondi (*right*) accretion rate estimates evaluated at two large radii: 100 pc (*top*) and 1 kpc (*bottom*). Solid diagonal line shows equality, dotted lines ± 1 dex. Colors denote gas fraction (inside each radius) as in Figure 5. The viscous and Bondi accretion rate estimates both fare poorly, with very large scatter and a large systematic offset for the Bondi estimate. In the simulations, the BH "accretion rate" is estimated using the inflow rate at < 0.1 pc in nuclear-scale re-simulations that have initial conditions taken from the results of intermediate-scale or galactic-scale simulations with the appropriate conditions at ~ 100 pc or ~ 1 kpc, respectively.

bient gas properties than the inflow rate produced by stellar gravitational torques and orbit crossing (eq. 19 & 36). To highlight this, Figure 8 shows the ratio of the true inflow rate through a given annulus R in our simulations to the viscous accretion rate (eq. 45) and the order of magnitude gravitational torque inflow rate from equation 19. Because we show the results as a function of scale, all simulations on various scales are shown together (but each is only shown at the well-resolved and self-consistently modeled radii appropriate to it). We also compare the simulation results to a generalization of the Bondi-Hoyle formula appropriate for larger radii in a galaxy, with $M_{BH} \rightarrow M_{enc}$, the total enclosed mass within a radius. This is strictly valid for inflow through an annulus R only if M_{enc} changes relatively slowly with R so that the spherical accretion sonic radius is at $> R$. We include it to emphasize how poorly the simple Bondi-Hoyle scaling does at describing the radial variation of $\dot{M}$ in galactic simulations (for discussion of how Bondi rates change when applied outside the Bondi radius, see Colpi et al. 2007; Volonteri et al. 2008). The results in Figure 8 are for randomly chosen times near the peak of activity in each of the simulations from Figures 5-6 (the re-simulations from Hopkins & Quataert 2010a). Note in particular that Figure 8 shows the results of three different sets of simulations, with the intermediate and nuclear-scale simulations using initial conditions drawn from the results of the larger-scale simulations; the boundaries between the different sets of simulations are shown with dotted lines. Runs shown also include systematic numerical resolution studies discussed in Hopkins & Quataert (2010a).

Figure 8 shows that the gravitational torque model works well at all radii, with $\lesssim 0.3$ dex scatter and without significant systematic deviations over the range of simulations we have carried out. By contrast, the viscous and Bondi accretion rate fare poorly by comparison. Not only is there factor of ~ 10 scatter, but there are very large systematics as a function of radius (and in some cases, as a function of model parameters such as gas fraction and/or ISM equation of state parameterization q_{eos}).

Figure 9 demonstrates a similar point by comparing different predictors of the BH accretion rate given the conditions in the galaxy at large radii. Specifically, we compare the viscous or Bondi-Hoyle accretion rate evaluated at a large radius R_0 (as would be done using such prescriptions as sub-grid models) to the true inflow rate near the accretion disk. The latter is taken here to be the inflow rate at < 0.1 pc in our re-simulations that use initial conditions drawn from the galactic-scale conditions at R_0 (as in Fig. 8).[3] For the viscous accretion rate, we use $\alpha_V = 0.1$, although we consider this normalization to be arbitrary. For the Bondi-Hoyle accretion rate, we use $\alpha_B = 1$, which assumes that the density and sound speed are relatively constant inside R_0.

Figure 9 shows that neither the Bondi nor viscous models predict the actual inflow rates at small radii. For the Bondi rate, there is essentially no correlation between the true and predicted inflow rates. For the viscous rate, $\alpha_V \sim 0.1$ gets the median $\dot{M}$ approximately correct, but there is factor $\gtrsim 10$ scatter and only a weak correlation between the true and predicted inflow rates.[4] The predicted inflow rate also depends systematically on the radius at which it is evaluated (as in Fig. 8). Variations in the sound speed c_s drive much of the horizontal scatter in the predicted $\dot{M}_{BH}$ in Figure 9, at a given true $\dot{M}_{BH}$. This is because the Bondi and viscous models both depend strongly on c_s (eqs. 44 & 45), whereas the true inflow rate is only a weak function of sound speed (at least for the stellar-dominated systems that are the particular focus here).

5.2 A Gravitational Torque Model

In this section, we use the analytic results from § 2 to derive a physically motivated estimate of the inflow rate $\dot{M}$ at a given radius in a galaxy, which approximates the numerical simulations described in § 4. Our goal is to develop a model of the inflow rate that can be extrapolated over a large range of spatial scales, both to provide physical insight and to approximate the BH accretion rate on small scales from properties estimated on galactic or ~ 100 pc scales (e.g., in cosmological simulations or semi-analytic models).

We assume that the galaxy properties are known at a radius R_0, outside the BH radius of influence. We also assume that the stellar disk and bulge have approximately power-law surface density profiles, i.e., $\Sigma_{d,*} \propto R^{-\eta_d}$ and $\Sigma_b \propto R^{-\eta_b}$. This is reasonably accurate in our simulations, so long as $R_0 < R_e$ (the galaxy effective radius), and it allows us to make analytic progress. We will typically consider $\eta_b \sim \eta_d$ as well, i.e., that the bulge and disk have similar density profiles.

As argued in §3, the gas in galaxies reaches a quasi steady-state in which its radial distribution is set by a competition between inflow to smaller radii and star formation (Thompson et al. 2005):

$$\frac{\partial}{\partial R} \dot{M}_{inflow} \sim 2\pi R \dot{\Sigma}_* . \tag{46}$$

---

[3] In a simple alpha-disk prescription, the inflow rate onto the outer edge of the disk corresponds to the actual accretion rate onto the BH. In practice, there may be a complex transition to this inner disk, for which some intermediate-scale model (e.g. Power et al. 2010) should be applied to map the efficiency of inflow through to the viscous disk.

[4] It may appear surprising that the local viscous estimate is even accurate to a factor of ~ 100. This is because the sound speed in our simulations is $\gg$ the true thermal speed of neutral molecular or atomic gas (which dominates the mass); using the latter for c_s would lead to a much smaller $\dot{M}_{viscous}$. Our viscous estimate is thus more akin to an estimate of the inflow rate produced by turbulence generated by stellar feedback or strictly local, sub-sonic gravitational instabilities (Gammie 2001; Lodato & Rice 2004; Durisen et al. 2007).

The qualitative behavior of solutions to equation (46) depends on the ratio of the star formation time t_* to the inflow time $t_{\rm inf}$. For $t_{\rm inf} \ll t_*$, the solution is $\dot{M}_{\rm inflow} \simeq {\rm const} \gg \dot{M}_*$, while if $t_{\rm inf} \gg t_*$, $\dot{M}_{\rm inflow}$ declines precipitously at smaller radii as gas is consumed by star formation. The observed star formation times in galaxies are $\sim 10 - 100$ dynamical times, which is comparable to the inflow time for $a \sim 0.01 - 0.1$. This is similar to the mode amplitudes in our simulations and thus we expect $t_{\rm inf} \sim t_*$. A reasonable analytic approximation to the numerical solutions of equation (46) in this limit is given by taking $\dot{M}_{\rm inflow} \simeq \dot{M}_* \simeq \pi R^2 \dot{\Sigma}_*$. More precisely, this approximate solution is valid when it itself implies $\dot{M}_{\rm inflow} \propto R^p$ with $p > 0$, so that $\dot{M}_{\rm inflow}$ declines at small radii (as it must in the correct solution). This is not guaranteed for all η_d, η_b, η_K, etc., but is satisfied for the parameters we utilize below. The exact normalization will depend on details of the star formation law beyond just the power-law scaling (e.g. small-scale structure not treated explicitly in the simulations), so we consider it somewhat uncertain.

In the potential of the galaxy, the dominant mode of angular momentum transport is an $m = 2$ bar mode. To estimate $\dot{M}_{\rm inflow}$, we use equation 36, with the WKB approximation for $|\Phi_1|$,

$$|\Phi_1| \approx |a|_{\rm max} \, 2\pi \, G\Sigma_d \, R \, |kR|^{-1}. \tag{47}$$

We will eventually take $kR \sim 1$ because the most important modes are global. Since we are focusing on small scales in the galaxy ($\lesssim 0.1 - 1$ kpc), we would like to know the saturation mode amplitude $|a|_{\rm max}$ appropriate for self-excited modes (rather than the mode amplitude induced directly by an external perturber). This is in general a complex function of non-linear processes such as heating by the mode, fragmentation, inflow, and momentum exchange with the halo. Hopkins & Quataert (2010a) show the saturation amplitude $|a|_{\rm max}$ in simulations on this scale as a function of various properties of the galaxy (their Figure 12); the results can be reasonably approximated by $|a|_{\rm max} \approx a_0 f_d$ where $f_d = \pi G\Sigma_d / \Omega^2 R \approx M_d(R)/M_{\rm enc}(R)$ and $a_0 \sim 0.3$. Given this, equation 46 can now be approximated as

$$\left(\frac{\Sigma_g}{\Sigma_K}\right)^{\eta_K - 1} = \alpha_d \left(\frac{\Sigma_d}{\Sigma_t}\right)^2 \Omega t_K \tag{48}$$

$$\alpha_d \equiv \frac{m a_0 |2\nu_{\Sigma_{d,*}} + \nu_{\Sigma_g} - 3\nu_\Omega| (2 + \nu_{\Sigma_t})^2 F(\zeta)}{4\pi (2 + \nu_\Omega) |kR|} \tag{49}$$

$$\nu_u \equiv \frac{\partial \ln u}{\partial \ln R} \tag{50}$$

where Σ_K and t_K are the parameters of the Schmidt-Kennicutt law (eq. 39), $\Sigma_t = \Sigma_d + \Sigma_b$, and we take $\Omega^2(R) \approx G M_{\rm enc}(<R) R^{-3}$. Of course, the dominant term need not be just a bar, but the important point is that the dimensional scaling is the same for each – in the linear regime, modes are independent so each mode enters as $\sim m a_m F(\zeta_m)$ and the relevant normalization reflects the sum over the spectrum of modes (in simulations, this typically does not differ from the single-mode normalization by more than a factor of a few).

Near where BH dominates the potential, an $m = 1$ mode often dominates the transport. Again, in real systems there can be a rich mode structure, but we focus on this case for convenience and tractability. For a power-law mass profile, $M(<R) = M_{\rm enc}(<R_0) (R/R_0)^{2-\eta}$, it is straightforward to solve for the radius $R_{\rm BH}$ at which the BH dominates the potential, i.e., where $M(<R_{\rm BH}) = M_{\rm BH}$. The inflow rate in towards the BH from this radius then follows from equation 36, but now using the coefficients evaluated for an $m = 1$ mode in a roughly Keplerian potential. Because the modes here are global, we assume that $\Phi_1(<R_{\rm BH}) = a G M_d(R_{\rm BH})/R_{\rm BH}$ in the BH potential (the local/WKB term is small

as $R \rightarrow 0$). The approximate analytic solution of equation (46) is again $\dot{M}_{\rm inflow}(R) \simeq \dot{M}_*(R)$, which implies $\Sigma_g(R < R_{\rm BH}) \propto R^{-\eta_{\rm BH}}$, with $\eta_{\rm BH} = 1/2 (\eta_K - 1)$. Specifically, in the limit of large gas supply, this leads to

$$\left(\frac{\Sigma_g}{\Sigma_K}\right)^{\eta_K - 1} = \alpha_{\rm BH} \sqrt{\frac{G M_{\rm BH}}{R_{\rm BH}^3}} \, t_K \left(\frac{R}{R_{\rm BH}}\right)^{-1/2} f_d(R_{\rm BH}) \tag{51}$$

$$\alpha_{\rm BH} = a_1 F(\zeta_1) \frac{3\eta_K - 4}{2\pi (\eta_K - 1)} \frac{2 + \nu_{\Sigma_K}}{2} \tag{52}$$

where $a_1 \approx 0.2$ is defined by $|a|_{\rm max} = |\Phi_1(R_{\rm BH})|/\Phi \approx a_1 f_d(R_{\rm BH})$ and the numerical value is based on the simulations of modes in the BH potential from Hopkins & Quataert (2010a). Together with the solution for $R_{\rm BH}$ using equation 48, this in turn gives the following BH inflow rate at a radius R:

$$\dot{M} = \tilde{\alpha} \, \Sigma_K R_0^2 \Omega(R_0) \left(\frac{R}{R_{\rm BH}}\right)^{\frac{3}{2} - \eta_{\rm BH}} \mu_{\rm BH}^\gamma \tag{53}$$

$$\times f_d^{\frac{3\eta_K (1 - \eta_d)}{2(\eta_K - 1)(2 - \eta_d)}} [\Omega(R_0) t_K]^{\frac{1}{\eta_K - 1}}$$

$$\Omega(R_0) \equiv \left(\frac{G M_{\rm enc}(R_0)}{R_0^3}\right)^{1/2} \tag{54}$$

$$\mu_{\rm BH} \equiv \frac{M_{\rm BH}}{M_d(R_0)} \tag{55}$$

$$\gamma \equiv \frac{-4 + (3 - \eta_d)\eta_K}{2(2 - \eta_d)(\eta_K - 1)} \tag{56}$$

$$\tilde{\alpha} \equiv (2 + \nu_{\Sigma_t}) a_1 F(\zeta_1) \alpha_{\rm BH}^{\frac{1}{\eta_K - 1}} \tag{57}$$

Equation 53 is an estimate of the BH inflow rate in the limit of large gas supply (because we have used the $\dot{M}_{\rm inflow} \sim \dot{M}_*$ solution to eq. 46). If the gas supply at $R_{\rm BH}$ is less than a critical amount the inflow rate will be smaller than given by equation 53. The critical gas mass $M_{\rm gas,c}$ which defines the large gas supply limit is found by integrating $\Sigma_{\rm gas}$ in equation 51 from $R = 0$ to $R = R_{\rm BH}$. Recall as well that the intermediate-scale physics described by equation 48 itself sets the amount of gas that reaches $R_{\rm BH}$. The gas mass supplied to this radius, $M_{\rm gas,bh}$, can be estimated by integrating $\Sigma_{\rm gas}$ as implied by equation 48 over the same range. For $M_{\rm gas,bh} \gtrsim M_{\rm gas,c}$, equation 53 for the BH inflow rate is reasonably accurate while for $M_{\rm gas,bh} \lesssim M_{\rm gas,c}$ it overestimates the inflow rate. To quantify the suppression in $\dot{M}$ for $M_{\rm gas,bh} \lesssim M_{\rm gas,c}$, one can solve equation 46 as a boundary value problem with a given gas mass $M_{\rm gas,bh}$ at $R_{\rm BH}$ (as, e.g., Thompson et al. 2005 did using somewhat different assumptions about the inflow and star formation physics). For low gas masses the solution is an inflow rate that is approximately independent of radius and $\propto M_{\rm gas,bh}$, implying that the suppression of $\dot{M}$ is given by

$$\dot{M} \rightarrow \dot{M} \left[1 + \frac{M_{\rm gas,c}}{M_{\rm gas,bh}}\right]^{-1} \tag{58}$$

$$\frac{M_{\rm gas,c}}{M_{\rm gas,bh}} = \alpha_{g1} \left(\frac{\Sigma_d}{\Sigma_t}\right)^{-\frac{1}{\eta_K - 1}} \tag{59}$$

$$\alpha_{g1} = \frac{4\eta_K - 5 - \eta_d}{4\eta_K - 5} \left(\frac{\alpha_{\rm BH}}{\alpha_d}\right)^{\frac{1}{\eta_K - 1}} \tag{60}$$

Following the same logic, a similar suppression factor in $\dot{M}$ should appear if the gas mass available at R_0 is too small. Specifically, the solution for the gas surface density profile between $R_{\rm BH}$ and R_0 in equation 48 is valid only in the limit of large mass supply. It requires that the available gas mass at R_0, $M_{\rm gas}(R_0)$, exceed a critical value $M_{\rm gas,i}$ that can be found by by integrating $\Sigma_{\rm gas}$ in

equation 48 from $R = 0$ to $R = R_0$. The resulting suppression factor is then

$$\dot{M} \rightarrow \dot{M} \left[1 + \frac{M_{\mathrm{gas,i}}}{M_{\mathrm{gas}}(R_0)} \right]^{-1} \quad (61)$$

$$M_{\mathrm{gas,i}} = \alpha_g \, \Sigma_K \, R_0^2 \, f_d^{\frac{2}{\eta_K - 1}} \left(\Omega(R_0) \, t_K \right)^{\frac{1}{\eta_K - 1}} \quad (62)$$

$$\alpha_g \equiv \frac{4\pi \, (\eta_K - 1)}{4 \, \eta_K - \eta_d - 5} \, \alpha_d^{\frac{1}{\eta_K - 1}} \quad (63)$$

We now combine terms and evaluate the numerical coefficients and exponents in equations 53 and 58. We adopt star formation parameters consistent with those measured by Bouché et al. (2007): $\Sigma_K = 10^8 \, M_\odot \, \mathrm{kpc}^{-2}$, $t_K = 0.41 \times 10^9 \, \mathrm{yr}^{-1}$, and $\eta_K = 7/4$. We find reasonable consistency with our simulation results and the profiles of observed cusp ellipticals by taking $\eta_b \approx \eta_d \approx 1/2$ over this radial range (see e.g. Lauer et al. 2007; Hopkins et al. 2009a). We assume that the $R > R_{\mathrm{BH}}$ mode is $m = 2$, while the $R < R_{\mathrm{BH}}$ mode is $m = 1$, and that both are global, so that $|kR|, \zeta \sim 1$. We emphasize that all of these are gross simplifications and shouldn't be taken to be universally true – however they provide a starting point for the order of magnitude normalization of our derived inflow rates.

Finally, equation (53) should be evaluated at the radius R_{acc} where accretion itself maintains stability to local gravitational perturbations, i.e., $Q \gg 1$. Inside this radius, star formation ceases, and $\dot{M} \simeq$ constant. The value of R_{acc} in fact depends on the mechanism of angular momentum transport. For local angular momentum transport, $R_{\mathrm{acc}} \sim 0.01$ pc, while for larger-scale torques, $R_{\mathrm{acc}} \sim 0.1$ pc (Goodman 2003). Our calculations favor the latter physics, so we take $R_{\mathrm{acc}} \sim 0.01 R_{\mathrm{BH}}$, which corresponds to $R_{\mathrm{acc}} \sim 0.1$ pc for typical parameters (it makes little difference, to within the scatter and accuracy of our formula, if we take R_{acc} to be a fixed physical radius instead).

With these assumptions we find that our analytic gravitational torque inflow model predicts a BH inflow rate of

$$\frac{\dot{M}_{\mathrm{BH,grav}}}{M_\odot \, \mathrm{yr}^{-1}} \approx \alpha(\eta_K) \, f_d^{7/6} \left(1 + 1.71 \, f_d^{-4/3} \right)^{-1} \quad (64)$$

$$\left(\frac{M_{\mathrm{BH}}}{10^8 \, M_\odot} \right)^{1/6} \left(\frac{M_d(R_0)}{10^9 \, M_\odot} \right)^1 \left(\frac{R_0}{100 \, \mathrm{pc}} \right)^{-3/2}$$

$$\left(\frac{R_{\mathrm{acc}}}{10^{-2} \, R_{\mathrm{BH}}} \right)^{5/6} \left[1 + \frac{f_0}{f_{\mathrm{gas}}(R_0)} \right]^{-1}$$

where $\alpha(\eta_K)$ is a dimensionless constant discussed below. In a simpler form, equation 64 becomes

$$\frac{\dot{M}_{\mathrm{BH,grav}}}{M_\odot \, \mathrm{yr}^{-1}} \approx \alpha(\eta_K) \, f_d^{5/2} \, M_{\mathrm{BH,8}}^{1/6} \, M_{d,9} \, R_{0,100}^{-3/2} \left(1 + f_0/f_{\mathrm{gas}} \right)^{-1} \quad (65)$$

with

$$f_0 \approx 0.31 \, f_d^2 \left(\frac{M_d(R_0)}{10^9 \, M_\odot} \right)^{-1/3} \quad (66)$$

$$f_{\mathrm{gas}}(R_0) \equiv M_{\mathrm{gas}}(R_0) / M_d(R_0) \quad (67)$$

Recall that f_d in equations 64 and 65 is the total (stellar and gas) disk mass fraction evaluated at R_0. Note also that these predictions for $\dot{M}_{\mathrm{BH,grav}}$ can, under a variety of circumstances, exceed the Eddington accretion rate. How much above Eddington the BH can in fact accrete depends on the uncertain physics of radiation-pressure dominated, super-Eddington accretion; much of the mass may in fact be unbound at radii smaller than we consider here. This is less likely to be a concern for $\dot{M}_{\mathrm{BH,grav}} \lesssim \dot{M}_{\mathrm{Edd}}$.
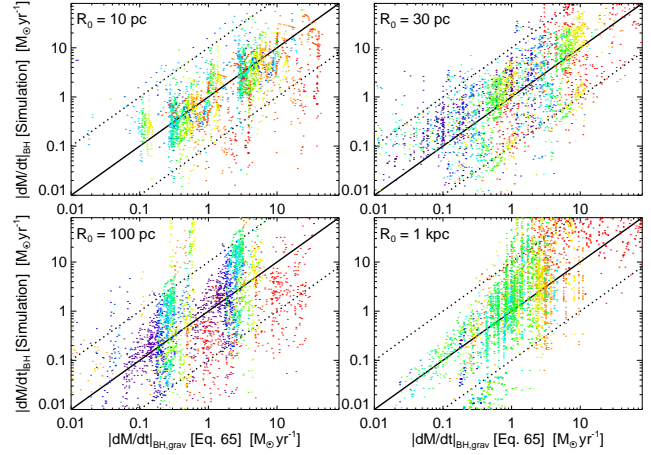

Figure 10. Comparison of the BH inflow rate (into < 0.1 pc) in our simulations with the predictions of our analytic model of inflow driven by stellar gravitational torques (eq. 65 with $\alpha(\eta_K) = 5$). In each panel, the properties used to infer $\dot{M}_{\mathrm{BH}}$ are measured at a different radius R_0, from 30 pc to 1 kpc (each from simulations modeling the appropriate scales). Colors denote gas fraction (inside the given radius) as in Figure 5. The analytic prediction works reasonably well and the agreement is much better than for the Bondi or viscous predictions at all radii (see Fig. 9). There remains, however, order-of-magnitude scatter – mostly owing to time variability in the BH inflow rate that should be included in models.

Despite the complex expressions leading to the final result for $\dot{M}_{\mathrm{BH,grav}}$, the power-law exponents in equations 64 and 65 are reasonably insensitive to the values of η_d and η_K over the plausible range ($0 \lesssim \eta_d \lesssim 1$, $1.4 \lesssim \eta_K \lesssim 1.75$). The normalization, however, is fairly sensitive, especially to η_K, ranging from $\approx 1 \, M_\odot$ yr to $\approx 18 \, M_\odot$ yr for $\eta_K = 7/4$ to $3/2$, respectively (for $\eta_d = 1/2$). This is one reason why we explicitly include the constant $\alpha(\eta_K) \sim 1 - 10$ in equations 64 and 65. The dependence on η_K is due to the fact that a larger η_K implies a higher star formation rate in dense gas and thus a lower BH inflow rate. Star formation on the small scales of interest here is uncertain – we choose to parameterize that in α so that the model has the advantage that it can be rescaled in a straightforward manner for many different models of star formation (via an appropriate choice of η_K or t_K). We have also made the oversimplified assumption of just one mode at each scale, when the real behavior reflects a sum over a mode spectrum; this does not change the dimensional scalings, but does enter the normalization α. The exact values we adopt here are therefore unlikely to be universally valid, but are convenient, observationally motivated, and match the choice in most of our simulations.

Figure 10 compares equation 65, with $\alpha(\eta_K) \approx 5$ and with the relevant quantities evaluated at different R_0 in the simulations, to the actual inflow rates into < 0.1 pc. There is, of course, very large scatter – even with full knowledge of the mass profile and mode amplitudes, the inflow rates are chaotic and time-variable. And unsurprisingly, the accuracy of equation 65 decreases for larger values of R_0. Moreover, because the profile shapes and mode structure do not exactly match what we have assumed at all radii, the appropriate normalization of $\dot{M}_{\mathrm{BH,grav}}$ depends weakly on R_0. In spite of its simplicity, however, equation 65 captures the broad trends in the BH inflow rate, and is a much more accurate approximation than either the viscous accretion rate or Bondi-Hoyle accretion rate (compare Figures 10 and 9).

As a sanity check on the predictions of equation 65, we briefly use it to estimate the inflow rate for two well studied systems:

the Milky Way and NGC 1068. For the Milky Way, which has $M_{BH} \simeq 4 \times 10^6 M_\odot$ (Ghez et al. 2008) and $f_{gas} \simeq 0.4$, $f_d \simeq 0.2$ and $M_{gas}(R_0) \simeq 3 \times 10^7 M_\odot$ at $R_0 \simeq 100$ pc (Serabyn & Morris 1996), equation 65 predicts $\dot{M}_{BH,grav} \sim 3 \times 10^{-3} M_\odot$ yr^{-1}. Although much larger than the current inflow rate onto Sgr A*, this is not unreasonable for a time averaged inflow rate given that BHs of $\sim 10^7 M_\odot$ are growing significantly in mass in the local universe (Heckman et al. 2004). For NGC 1068, a nearby Type 2 Seyfert with $M_{BH} \simeq 1.5 \times 10^7 M_\odot$ (Greenhill & Gwinn 1997) and $L \sim 10^{45}$ erg s$^{-1} \sim L_{Edd}$ (Pier et al. 1994), the gas mass within $R_0 \simeq 100$ pc is $\sim 3 \times 10^7 M_\odot$ while the total dynamical mass is $\sim 3 \times 10^8 M_\odot$ (Helfer & Blitz 1995). The total disk mass within this radius is uncertain but Davies et al. (2007) find evidence for a stellar disk with $\sim 10^8 M_\odot$ within ~ 100 pc (or less). Adopting $f_{gas} \sim 1/3$ and $f_d \sim 0.4$, equation 65 predicts $\dot{M}_{BH,grav} \sim 0.03 M_\odot$ yr^{-1} comparable to the Eddington accretion rate $\sim 0.1 M_\odot$ yr^{-1} for NGC 1068. There is, of course, no reason to expect that our estimate should accurately predict the inflow rate in a given system at a given time (as indeed it does not for the Milky Way!). It is, however, reassuring that our estimates of $\dot{M}$ for the Milky Way and NGC 1068 are observationally plausible in a time-averaged sense.

It is worth highlighting how and why our inflow rate prediction (eq. 65) depends on several key physical parameters:

- f_d: The inflow rate decreases rapidly with decreasing disk fraction (i.e., increasing bulge to disk ratio). This is expected for inflow driven by global gravitational instability in the disk.
- M_{BH}: The absolute inflow rate is not a strong function of M_{BH}. In units of the Eddington accretion rate, the inflow rate actually decreases with increasing BH mass ($\sim M_{BH}^{-5/6}$ for $\eta_d = 1/2$). This is because on small scales, the BH plays the same role as the bulge, suppressing global modes within R_{BH}.
- M_d: $\dot{M}_{BH,grav}$ is linear in M_d, the *total* (gas+stellar) disk mass.
- f_{gas}: Provided that sufficient gas is available ($f_{gas} \gtrsim f_0$), the inflow rates are nearly independent of the gas supply, because the gas densities at small scales are set by an equilibrium between inflow and star formation, which self-adjusts to a particular inflow rate. Of course, at very high f_{gas}, our assumption of stellar-dominated disks breaks down, and it is not clear how robust the extrapolated model will be.
- $\dot{M}_*$: The star formation rate enters into the expression for $\dot{M}_{BH,grav}$ via the values of t_K, Σ_K, and η_K from the Schmidt-Kennicut law (see eq. 53 for the precise dependencies). A lower star formation efficiency (corresponding to larger t_K and/or smaller η_K) at small radii in gas-rich galactic nuclei, as several authors have suggested (Thompson et al. 2005; Begelman & Shlosman 2009; Larson 2009), would significantly increase the final BH accretion rate, making it easier to grow a central massive BH.
- c_s: In our model, the inflow rate is independent of the sound speed c_s and other thermal properties of the gas. This is because the torques and inflow rate are determined by supersonic non-axisymmetric motions induced by global gravitational instabilities. Again, caution is needed when systems have very large f_{gas} and fragmentation can compete with global torques.

Evaluating equation 65 requires the input parameters M_{BH}, $M_{gas}(R_0)$, $M_d(R_0)$, $M_b(R_0)$ and a choice of R_0 (ideally the smallest resolved radii, since Fig. 10 shows that the analytic approximations here are best at smaller radii $\lesssim 100$ pc – but they are still far superior to the Bondi and viscous models even at $\sim$ kpc). In analytic and semi-analytic models, these are all clearly-defined, and implementing equation 65 is straightforward. In numerical simulations, it is less clear how to define the bulge and disk separately in an on-the-fly manner. One possibility that may be sufficient at the level of accuracy of our model is to define the disk mass fraction via the total angular momentum of the material within R_0 of the BH.

As noted above, there is considerable variability on a range of timescales in our simulations of AGN fueling – see Hopkins & Quataert (2010a) for a more extensive discussion. Very roughly, we find approximately equal power in fluctuations on each dynamical time resolved in the simulation. Recently, Levine et al. (2010) show this explicitly for timescales from $\gtrsim 10^7$ to ~ 100 yr (see their Figures 3-5) using AMR simulations that cover a range of radii similar to our simulations. Given this variability, it is unlikely that *any* BH accretion rate model could predict inflow rates on small scales given the properties of a galaxy on large scales with much better than order-of-magnitude accuracy (as in Fig. 10). Moreover, this variability can have important consequences for the effects of AGN feedback or the ionization of the intergalactic medium, to cite just two examples. We thus believe that it is important to include variability when modeling BH growth and AGN feedback.[5]

5.3 Implications of our Gravitational Torque Model

Having derived the new inflow rate predictor in equation 65 and shown that its functional form is significantly different from the Bondi or viscous accretion rate estimates, we briefly speculate on conditions under which we do or do not expect this to make a significant difference relative to previous results. Because the problem of black hole fueling with feedback is highly nonlinear, it is entirely possible that the results of applying equation 65 in numerical simulations could be more subtle than we anticipate. Thus it will be critical to explore this numerically as well.

We anticipate that our new fueling model may only have limited effects in the following circumstances:

- Eddington-Limited Accretion: In e.g. simulations of galaxy mergers, the AGN passes through a significant phase in which very large inflows from galactic scales lead to a large gas reservoir in the central ~ 100 pc. The implied accretion rates for a range of models are comparable to, or in excess, the Eddington rate (Di Matteo et al. 2005; Debuhr et al. 2010). This suggests that the Eddington limit is the rate-limiting factor in many cases, not the rate of mass supply.
- Feedback-Regulated Accretion: Once BHs reach a certain mass, models that include BH feedback often lead to self-regulated BH growth. In this limit, both simulations and analytic arguments suggest that the accretion rate is determined by the balance between feedback and inflow, independent of the precise functional form of the accretion rate model (Debuhr et al. 2011). Because the BH gains most of its mass during this phase, we suspect that the new accretion rate model proposed here will not significantly change previous inferences about the conditions required to reproduce the $M_{BH} - \sigma$ relation and other BH-galaxy correlations.

---

[5] For example, following Levine et al. (2010), in semi-analytic models or galactic-scale simulations one could approximate the variability on timescales smaller than some resolved timescale $t_0 \sim \Omega_0(R_0)$, and down to a minimum variability timescale dt, using a Fourier series

$$\dot{M} = \langle \dot{M} \rangle \prod_{i=0}^{N} \exp \left\{ \frac{\sigma_0 \ln 10}{\sqrt{N/2}} f_i \cos \left[\frac{\Omega_0 t}{(\Omega_0 dt)^{i/N}} + 2\pi \delta_i \right] \right\} \quad (68)$$

where $\sigma_0 \sim 0.5 - 0.6$ is the amplitude of variations (motivated by the simulations) and the f_i and δ_i are a Gaussian random numbers drawn with dispersion of unity and uniform random number from $0 - 1$, respectively; $N \gtrsim 10$ appears sufficient for a reasonable approximation of the variability in the simulations.

There are, however, other contexts in which we suspect that the more physical treatment of BH fueling presented here will have a significant effect on the results of analytic models and galactic-scale simulations:

• AGN Lightcurves: Most BHs, most of the time, accrete below the Eddington rate. In this limit, the predicted AGN accretion rate is more sensitive to the details of the accretion rate model. As a result, our new accretion rate predictions may significantly affect the AGN lightcurves and the fraction of time a BH spends at a given accretion rate. As emphasized by Hopkins et al. (2005), this will in turn affect the interpretation of the AGN luminosity function.

• Early BH Growth: The Bondi accretion rate is $\propto M_{\rm BH}^2$, which suppresses accretion onto low mass BH seeds at high redshift. By contrast, the model presented here depends only weakly on BH mass – low BH mass in fact allows global instabilities to form more readily and thus should promote rapid seed growth.

• BH Growth Versus Galaxy Morphology: Our accretion model includes an explicit, strong dependence on the nuclear morphology of the galaxy. BH growth via secular processes will be relatively more efficient in disk-dominated galaxies, but suppressed in bulge-dominated galaxies, potentially requiring more violent events (mergers) to grow BHs. Because the morphology as a function of radius is also important, the bulge structural properties are also relevant for the resulting BH accretion rate: e.g., low mass but sufficiently compact bulges can suppress accretion on small scales, despite allowing inflow to the central ∼kpc. This could introduce differences between the predicted accretion rates in different types of bulges (e.g. classical or pseudo-bulges). The dependence of the accretion rate on galaxy morphology may end up being one of the most important differences between the predictions of our model and previous results.

6   DISCUSSION AND CONCLUSIONS

The dominant mechanism generating significant gas inflows from ∼ 10 kpc to ∼ 100 pc in galaxies has been extensively studied using numerical simulations over the past two decades: a perturbation to the galaxy (e.g., a merger or disk instability) generates a non-axisymmetric structure. If the perturbation is sufficiently strong in the collisionless component of the system, it drives a corresponding perturbation in the gaseous component. This, in turn, forces the gas to shock, which dissipates energy and relative motion, allowing for net angular momentum loss and inflow. We have recently demonstrated that qualitatively similar physics can operate from ∼ 0.1 − 10 pc in galactic nuclei if they also have a dominant collisionless (stellar) component, although the relevant non-axisymmetric mode is preferentially $m = 1$ instead of $m = 2$ as on larger scales (Hopkins & Quataert 2010a). The critical physics of stellar-induced shocks in gas has thus been seen in simulations of galaxy-galaxy mergers, unstable galactic disks, and secondary and tertiary gravitational instabilities in galactic nuclei (Noguchi 1988; Barnes & Hernquist 1996; Barnes 1998; Berentzen et al. 2007; Hopkins et al. 2009b; Hopkins & Quataert 2010a). In this paper, we have developed an analytic framework to model this physics and to describe how gravitational perturbations redistribute angular momentum, and drive shocks and gas inflow, in galaxies. We have also tested these models against a large suite of numerical simulations.

Dissipation and the co-existence of collisional (gas) and collisionless (stars) disky components are critical for generating large gaseous inflow rates. Dissipationless components can re-distribute angular momentum via resonances (Lynden-Bell & Kalnajs 1972; Goldreich & Tremaine 1979; Hernquist & Weinberg 1992; Athanassoula 2003; Weinberg & Katz 2007) but these are much less efficient than gravitationally-induced shocks, with inflow rates smaller by a factor of $\gtrsim 10 - 100$. And without a significant fraction of the mass in a collisionless component that drives the gas into shocks, the gas orbits would self-adjust to eliminate any long-lived strong shocks.

The canonical criterion for gas orbits to shock in the presence of a non-axisymmetric potential perturbation is that $|\zeta| = |dR_1/dR_0| > 1$, where R_1 is the perturbation to a circular orbit initially at radius R_0 due to the non-axisymmetric potential (eq. 20). This criterion is, however, only a sufficient condition for shocks, not a necessary one. In the case of shocks near spiral arms or bars in a self-gravitating disk, the traditional condition on $|\zeta|$ is often adequate, since $|\zeta| \gg 1$ for moderate-amplitude perturbations. But this is not always the case, particularly for disks in a quasi-Keplerian potential near a BH. We have demonstrated that even when $|\zeta| \lesssim 1$, a stellar perturbation can produce shocks in a gaseous disk with a finite sound speed (eq. 29-30). Physically, the key point is that when the non-axisymmetric potential is stellar dominated, gas "collisions" can happen not just at strict orbit crossings, but at a finite separation between streamlines. The key requirements are that the amplitude of the non-axisymmetric stellar perturbation exceed the critical value given in eq. 30, and that a reasonable fraction of the mass is in a disky collisionless component. We have not quantified in detail how large the mass fraction in the collisionless component must be, but simple force considerations suggest $\gtrsim 50\%$.

Given the presence of shocks in the gas, we have derived the resulting angular momentum loss/gain and the gas inflow/outflow rates in two distinct limits: very strong shocks ($|\zeta| \gg 1$) and marginal orbit crossings ($|\zeta| \lesssim 1$) (eq. 36). This calculation includes a proper treatment of the phase function for angular momentum exchange between the gas and stars, i.e., when the net torque produces gas inflow and when it produces outflow. For typical conditions in galaxies, we find flow towards an inner Lindlad resonance inside of co-rotation and towards an outer Lindlad resonance outside of co-rotation. However, the behavior can be more complex depending on the details of the system.

One of our key results is that the gas inflow produced by gravitationally-induced shocks is robust and rapid, with

$$\dot{M} \sim \frac{|a| M_{\rm gas}}{t_{\rm dyn}} \sim |a| M_{\rm gas} \Omega \qquad (69)$$

where $|a|$ is the fractional non-axisymmetric perturbation to the potential (eq. 19 & 36). Note several key facts about this scaling. First, $\dot{M}$ is *linear* in $|a|$. In contrast, when shocks do not occur, angular momentum transport by non-axisymmetric perturbations is at least second-order in $|a|$ (§2.2). Moreover, this second-order transport is also suppressed by powers of kR and m where k (m) is the radial (azimuthal) wavenumber of the non-axisymmetric mode (see, e.g., Kalnajs 1971's exact solution for the logarithmic spiral; eq 18). Given these differences, we find inflow rates that are a factor of ∼ 10 − 100 larger than would traditionally be predicted by linear theory (for a given set of gas properties).

A second key property of equation 69 is that the inflow rate is independent of the sound speed of the gas c_s, i.e., of the gas thermodynamics. Physically, this is because the angular momentum transport is governed by the rate at which gas orbits are forced to shock via the non-axisymmetric stellar potential. Because the relevant gas motions are by definition supersonic and induced by a collisionless component, the precise gas thermodynamics is not important. This

property of inflow produced by gravitationally-induced shocks is fundamentally different from angular momentum transport by local viscous stresses ("α") or by linear spiral waves, both of which depend on c_s.[6] The lack of an explicit dependence on the gas thermodynamics, which is particularly uncertain in a multi-phase interstellar medium with stellar feedback, points to the robustness of our predicted inflow rates in this regime. One potential subtlety is that our analytic work assumes that a significant fraction of the ISM consists of relatively diffuse gas, rather than gas bound up in, e.g., star clusters; this is also true to some extent in our numerical work (see §4.1 for a discussion of this point). This assumption is motivated by the observed low star formation efficiency in dense gas and the small fraction of a molecular cloud's mass that is converted into stars before the cloud is disrupted (e.g., Krumholz & Tan 2007). But further numerical work including a more detailed treatment of stellar feedback and a multi-phase ISM would shed considerable light on whether a highly inhomogeneous ISM qualitatively modifies the physics of gas inflow in galaxies. It is well-known, for example, that mode development depends significantly on gas thermodynamics in pure gas disks (e.g. Gammie 2001; Nayakshin et al. 2007; Cossins et al. 2009). We suspect that the physics described here will be applicable to any collisionless plus collisional disk system, provided that the post-shock cooling time of the gas is short compared to the local dynamical time and the collisionless component is gravitationally dominant (well-satisfied in galaxies at the radii of interest).

Another way to appreciate the importance of equation 69 is to compare the resulting inflow time of gas in galaxies ($\sim |a|^{-1} t_{\rm dyn}$) to the star formation timescale. The latter is observed to be $\sim 10 - 100\ t_{\rm dyn}$ (e.g., Bouché et al. 2007). Since the non-axisymmetric mode amplitudes produced by mergers or instabilities of self-gravitating disks can readily reach $|a| \sim 0.01 - 1$ (Fig. 12 of Hopkins & Quataert 2010a), equation 69 implies that the gas inflow time can be comparable to or shorter than the timescale for the gas to turn into stars. This is the fundamental reason that we find BH inflow rates $\sim 1 - 10 M_\odot\ {\rm yr}^{-1}$ in our simulations (sufficient to fuel luminous quasars) without overproducing star formation in galactic nuclei. Similar inflow rates have also been found in independent simulations by Levine et al. (2010). By contrast, a number of previous calculations (e.g., Goodman 2003; Sirko & Goodman 2003; Thompson et al. 2005; Nayakshin & King 2007) assumed either a local α model or linear transport by spiral waves, neither of which accurately describe our analytic or numerical results. Because of this incorrect treatment of angular momentum transport, these previous calculations also significantly overestimated the inflow time of gas in galaxies, and thus overestimated the fraction of the inflowing gas that turns into stars. In our simulations, we find that star formation and inflow reach an approximate equilibrium, in which the inflow rate is comparable to the local star formation rate (except at sufficiently small radii, where the inflow rate becomes independent of radius). This balance determines the quasi-steady gas surface density as a function of radius (§ 3), and is in good agreement with the numerical results from $\sim 10^{-3} - 10$ kpc (e.g., Figs 2 & 3). Of course, star formation in these models is still "sub-grid" and pegged to an empirical law, so avoiding runaway local collapse may still require processes such as stellar feedback.

For the purposes of analytic or semi-analytic calculations and galactic or cosmological numerical simulations, it would be use-

ful to have an estimate of the BH inflow rate based only on galaxy properties at much larger radii (generally > 100 pc). Calculations in the literature often assume simple models such as Eddington-limited accretion for a fixed timescale, spherical Bondi accretion (Springel et al. 2005; Di Matteo et al. 2005), local viscous transport (e.g., Sirko & Goodman 2003; Debuhr et al. 2010), or linear transport by spiral waves (e.g., Thompson et al. 2005). None of these models, however, accounts for the key physics of gravitationally-induced shocks in the gas. We have explicitly demonstrated that both spherical accretion and local viscous transport are relatively poor approximations to the inflow seen in our numerical simulations (Figs. 8 & 9).

Using our analytic models of gravitationally-induced gas inflow we have developed a more accurate predictor of the BH accretion rate given the large-scale properties of a galaxy (eqs 64-67). This model is highly simplified, and clearly far from exact, but it includes well-motivated dependencies for some of the key parameters of the problem. For example, the inflow rate is strongly suppressed for larger bulge-to-disk ratios, as is generally the case for global gravitational instabilities. In addition, the inflow rate is essentially independent of the gas sound speed so long as cooling is efficient and most of the gas resides in a rotationally-supported disk. By again comparing to our suite of numerical simulations, we have demonstrated that this new BH inflow predictor represents a major improvement relative to previous models (Fig. 10); see §5.3 for a brief discussion of some of the implications of this result. However, even this new BH predictor has a factor of ~ 10 scatter when compared to the simulations. This does not appear to be due to any systematic dependence on galaxy structural properties, gas fraction, etc. that is not already captured in our scalings. Rather, the scatter is a consequence of physical variability in the inflow rate for a given set of conditions in the galaxy at large radii. Indeed, because gravitational instabilities operate on all scales, and because the systems of interest are chaotic, we find that there is variability on essentially all timescales (see also Levine et al. 2010). We believe that it is important to include this variability in future models of BH growth and AGN feedback.

A number of important regimes of gas inflow require significant further study. We have assumed throughout this paper that the non-axisymmetric mode amplitudes are large enough to produce shocks at nearly all radii. This is indeed borne out by our (and others') simulations of galaxy-galaxy mergers and gas-rich galactic nuclei, but it is by no means clear that this will always be the case. For low mode amplitudes, we do not expect shocks and so the angular momentum transport mechanisms studied in this paper will not be present. Understanding which processes dominate inflow at lower inflow rates will be of considerable interest for models of low-luminosity AGN and weak enhancements in nuclear star formation.

The regime in which galaxies are gas-dominated is also of considerable interest. In this limit, it is unlikely that strong orbit crossings and/or coherent long-lived shocks will occur, if the mode remains linear. Generically, we expect this to decrease the efficiency of gas inflow. This has in fact been seen and studied in detail in the context of galaxy-galaxy mergers (e.g., Hopkins et al. 2009d). Of course, if strongly non-linear modes can be sustained, then our assumptions are probably not applicable and there can in principle be a wide range of inflow rates – a pure radial inflow at the free-fall velocity is a valid non-linear configuration! And in very gas-rich systems, local fragmentation may be much more efficient, generating an inhomogeneous ISM that may provide mechanisms for angular momentum transport that we have not considered in

---

[6] More precisely, transport by linear spiral waves only depends on c_s indirectly, if the system maintains $Q \sim 1$; see eq. 18.

this paper. Understanding this physics on smaller scales in galactic nuclei is particularly important for high-redshift galaxy formation and the formation of the first BHs, where it has been posited that they might form in direct collapse of primordial gas clouds (Begelman & Shlosman 2009, and references therein). In our simulations of gas dominated systems to date (Hopkins & Quataert 2010a), we find that the gas inflow from ~ 10 pc to ~ 0.1 pc is significantly more efficient after some fraction of the gas has turned into stars. This is qualitatively consistent with the results from galaxy merger simulations, but needs to be studied in more detail (for comparison, see e.g. Agertz et al. 2009; Teyssier et al. 2010).

ACKNOWLEDGMENTS

REFERENCES

Adams, F. C., Ruden, S. P., & Shu, F. H. 1989, ApJ, 347, 959
Agertz, O., Teyssier, R., & Moore, B. 2009, MNRAS, 397, L64
Athanassoula, E. 1992a, MNRAS, 259, 328
—. 1992b, MNRAS, 259, 345
—. 2003, MNRAS, 341, 1179
Balbus, S. A., & Hawley, J. F. 1998, Reviews of Modern Physics, 70, 1
Barnes, J. E. 1998, in Saas-Fee Advanced Course 26: Galaxies: Interactions and Induced Star Formation; Springer-Verlag Berlin/Heidelberg, ed. R. C. Kennicutt, Jr., F. Schweizer, J. E. Barnes, D. Friedli, L. Martinet, & D. Pfenniger, 275–+
Barnes, J. E., & Hernquist, L. 1996, ApJ, 471, 115
Barnes, J. E., & Hernquist, L. E. 1991, ApJL, 370, L65
Begelman, M. C., & Shlosman, I. 2009, ApJL, 702, L5
Berentzen, I., Shlosman, I., Martinez-Valpuesta, I., & Heller, C. H. 2007, ApJ, 666, 189
Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846
Binney, J., & Tremaine, S. 1987, Galactic dynamics (Princeton, NJ, Princeton University Press, 1987)
Block, D. L., Puerari, I., Knapen, J. H., Elmegreen, B. G., Buta, R., Stedman, S., & Elmegreen, D. M. 2001, Astronomy and Astrophysics, 375, 761
Bouché, N., et al. 2007, ApJ, 671, 303
Bournaud, F., Combes, F., & Semelin, B. 2005, MNRAS, 364, L18
Bournaud, F., Elmegreen, B. G., & Elmegreen, D. M. 2007, The Astrophysical Journal, 670, 237
Bournaud, F., Jog, C. J., & Combes, F. 2005, A&A, 437, 69
Bryant, P. M., & Scoville, N. Z. 1999, AJ, 117, 2632
Carlberg, R. G. 1986, ApJ, 310, 593
Colpi, M., Callegari, S., Dotti, M., Kazantzidis, S., & Mayer, L. 2007, in American Institute of Physics Conference Series, Vol. 924, The Multicolored Landscape of Compact Objects and Their Explosive Origins, ed. T. di Salvo, G. L. Israel, L. Piersant L. Burderi G. Matt A. Tornambe & M. T. Menna, 705–714
Combes, F., et al. 2004, A&A, 414, 857
Cossins, P., Lodato, G., & Clarke, C. J. 2009, MNRAS, 393, 1157

Cox, T. J., Dutta, S. N., Di Matteo, T., Hernquist, L., Hopkins, P. F., Robertson, B., & Springel, V. 2006, ApJ, 650, 791
Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, MNRAS, 384, 386
Dasyra, K. M., et al. 2007, ApJ, 657, 102
Davies, R. I., Sánchez, F. M., Genzel, R., Tacconi, L. J., Hicks, E. K. S., Friedrich, S., & Sternberg, A. 2007, ApJ, 671, 1388
Debuhr, J., Quataert, E., & Ma, C. 2011, MNRAS, 412, 1341
Debuhr, J., Quataert, E., Ma, C., & Hopkins, P. 2010, MNRAS, 406, L55
Di Matteo, T., Springel, V., & Hernquist, L. 2005, Nature, 433, 604
Dotti, M., Ruszkowski, M., Paredi, L., Colpi, M., Volonteri, M., & Haardt, F. 2009, MNRAS, 396, 1640
Durbala, A., Buta, R., Sulentic, J. W., & Verdes-Montenegro, L. 2009, MNRAS, 397, 1756
Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., & Rice, W. K. M. 2007, Protostars and Planets V (B. Reipurth, D. Jewitt, and K. Keil (eds.), University of Arizona Press, Tucson), 951, 607
Eliche-Moral, M. C., González-García, A. C., Balcells, M., Aguerri, J. A. L., Gallego, J., & Zamorano, J. 2008, in Astronomical Society of the Pacific Conference Series, Vol. 396, Formation and Evolution of Galaxy Disks, Astronomical Society of the Pacific Conference Series, ed. J. G. Funes & E. M. Corsini, 359–+
Escala, A. 2007, ApJ, 671, 1264
Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2004, ApJ, 607, 765
Gammie, C. F. 2001, ApJ, 553, 174
García-Burillo, S., Combes, F., Schinnerer, E., Boone, F., & Hunt, L. K. 2005, Astronomy and Astrophysics, 441, 1011
Genzel, R., et al. 2008, ApJ, 687, 59
Ghez, A. M., et al. 2008, ApJ, 689, 1044
Goldreich, P., & Lynden-Bell, D. 1965a, MNRAS, 130, 97
—. 1965b, MNRAS, 130, 125
Goldreich, P., & Tremaine, S. 1978, ApJ, 222, 850
—. 1979, ApJ, 233, 857
—. 1980, ApJ, 241, 425
Goodman, J. 2003, MNRAS, 339, 937
Greene, J. E., Zakamska, N. L., Liu, X., Barth, A. J., & Ho, L. C. 2009, ApJ, 702, 441
Greenhill, L. J., & Gwinn, C. R. 1997, Astrophysics and Space Science, 248, 261
Gunn, J. E. 1987, in Nearly Normal Galaxies. From the Planck Time to the Present (New York: Springer-Verlag), ed. S. M. Faber, 455–464
Haan, S., Schinnerer, E., Emsellem, E., García-Burillo, S., Combes, F., Mundell, C. G., & Rix, H.-W. 2009, The Astrophysical Journal, 692, 1623
Heckman, T. M., Kauffmann, G., Brinchmann, J., Charlot, S., Tremonti, C., & White, S. D. M. 2004, ApJ, 613, 109
Helfer, T. T., & Blitz, L. 1995, ApJ, 450, 90
Hernquist, L. 1989, Nature, 340, 687
—. 1990, ApJ, 356, 359
—. 1993, ApJ, 404, 717
Hernquist, L., & Mihos, J. C. 1995, ApJ, 448, 41
Hernquist, L., Spergel, D. N., & Heyl, J. S. 1993, ApJ, 416, 415
Hernquist, L., & Weinberg, M. D. 1992, ApJ, 400, 80
Hibbard, J. E., & Yun, M. S. 1999, ApJL, 522, L93
Hicks, E. K. S., Davies, R. I., Malkan, M. A., Genzel, R., Tacconi, L. J., Sánchez, F. M., & Sternberg, A. 2009, ApJ, 696, 448

Hopkins, P. F., Cox, T. J., Dutta, S. N., Hernquist, L., Kormendy, J., & Lauer, T. R. 2009a, ApJS, 181, 135

Hopkins, P. F., Cox, T. J., Younger, J. D., & Hernquist, L. 2009b, ApJ, 691, 1168

Hopkins, P. F., & Hernquist, L. 2006, ApJS, 166, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., Di Matteo, T., Robertson, B., & Springel, V. 2005, ApJ, 630, 716

Hopkins, P. F., Lauer, T. R., Cox, T. J., Hernquist, L., & Kormendy, J. 2009c, ApJS, 181, 486

Hopkins, P. F., & Quataert, E. 2010a, MNRAS, 407, 1529

—. 2010b, MNRAS, 405, L41

Hopkins, P. F., et al. 2009d, MNRAS, 397, 802

Kalnajs, A. J. 1971, ApJ, 166, 275

Kawakatu, N., & Wada, K. 2008, ApJ, 681, 73

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

Klessen, R. S., Spaans, M., & Jappsen, A. 2007, MNRAS, 374, L29

Knapen, J. H., Pérez-Ramírez, D., & Laine, S. 2002, Monthly Notice of the Royal Astronomical Society, 337, 808

Knapen, J. H., Shlosman, I., & Peletier, R. F. 2000, ApJ, 529, 93

Kormendy, J. 1989, ApJL, 342, L63

Krumholz, M. R., & Tan, J. C. 2007, ApJ, 654, 304

Laine, S., Shlosman, I., Knapen, J. H., & Peletier, R. F. 2002, The Astrophysical Journal, 567, 97

Larson, R. B. 2009, Nature Physics, 6, 96

Lau, Y. Y., & Bertin, G. 1978, ApJ, 226, 508

Lauer, T. R., et al. 1993, AJ, 106, 1436

—. 2007, ApJ, 664, 226

Levine, R., Gnedin, N. Y., & Hamilton, A. J. S. 2010, ApJ, 716, 1386

Levine, R., Gnedin, N. Y., Hamilton, A. J. S., & Kravtsov, A. V. 2008, ApJ, 678, 154

Lin, C. C., Yuan, C., & Shu, F. H. 1969, ApJ, 155, 721

Lodato, G., & Rice, W. K. M. 2004, MNRAS, 351, 630

Lynden-Bell, D., & Kalnajs, A. J. 1972, MNRAS, 157, 1

Martini, P., & Pogge, R. W. 1999, Astron. J., 118, 2646

Mayer, L., Kazantzidis, S., Escala, A., & Callegari, S. 2010, Nature, 466, 1082

Mayer, L., Kazantzidis, S., Madau, P., Colpi, M., Quinn, T., & Wadsley, J. 2007, Science, 316, 1874

Mo, H. J., Mao, S., & White, S. D. M. 1998, MNRAS, 295, 319

Murray, N., Quataert, E., & Thompson, T. A. 2010, ApJ, 709, 191

Nayakshin, S., Cuadra, J., & Springel, V. 2007, MNRAS, 379, 21

Nayakshin, S., & King, A. 2007, MNRAS, in press, arXiv:0705.1686

Noguchi, M. 1988, A&A, 203, 259

O'Shea, B. W., Nagamine, K., Springel, V., Hernquist, L., & Norman, M. L. 2005, ApJS, 160, 1

Ostriker, E. C., Shu, F. H., & Adams, F. C. 1992, ApJ, 399, 192

Ostriker, J. P. 1980, Comments on Astrophysics, 8, 177

Papaloizou, J., & Pringle, J. E. 1977, MNRAS, 181, 441

Peletier, R. F., et al. 1999, ApJS, 125, 363

Pier, E. A., Antonucci, R., Hurt, T., Kriss, G., & Krolik, J. 1994, ApJ, 428, 124

Power, C., Nayakshin, S., & King, A. 2010, MNRAS, in press, arXiv:1003.0605

Quillen, A. C., Frogel, J. A., Kenney, J. D. P., Pogge, R. W., & Depoy, D. L. 1995, ApJ, 441, 549

Roberts, W. W., Huntley, J. M., & van Albada, G. D. 1979, Astrophysical Journal, 233, 67, a&AA ID. AAA026.151.031

Robertson, B., Cox, T. J., Hernquist, L., Franx, M., Hopkins, P. F., Martini, P., & Springel, V. 2006a, ApJ, 641, 21

Robertson, B., Hernquist, L., Cox, T. J., Di Matteo, T., Hopkins, P. F., Martini, P., & Springel, V. 2006b, ApJ, 641, 90

Sakamoto, K., Okumura, S. K., Ishizuki, S., & Scoville, N. Z. 1999, The Astrophysical Journal, 525, 691

Sanders, D. B., Soifer, B. T., Elias, J. H., Madore, B. F., Matthews, K., Neugebauer, G., & Scoville, N. Z. 1988a, ApJ, 325, 74

Sanders, D. B., Soifer, B. T., Elias, J. H., Neugebauer, G., & Matthews, K. 1988b, ApJL, 328, L35

Schinnerer, E., Böker, T., Emsellem, E., & Lisenfeld, U. 2006, ApJ, 649, 181

Serabyn, E., & Morris, M. 1996, Nature, 382, 602

Shlosman, I., & Begelman, M. C. 1989, ApJ, 341, 685

Shlosman, I., Frank, J., & Begelman, M. C. 1989, Nature, 338, 45

Shu, F. H., Tremaine, S., Adams, F. C., & Ruden, S. P. 1990, ApJ, 358, 495

Sirko, E., & Goodman, J. 2003, MNRAS, 341, 501

Solomon, P. M., Downes, D., Radford, S. J. E., & Barrett, J. W. 1997, ApJ, 478, 144

Spaans, M., & Silk, J. 2005, ApJ, 626, 644

Springel, V. 2005, MNRAS, 364, 1105

Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776

Springel, V., & Hernquist, L. 2002, MNRAS, 333, 649

—. 2003, MNRAS, 339, 289

Tacconi, L. J., Genzel, R., Tecza, M., Gallimore, J. F., Downes, D., & Scoville, N. Z. 1999, ApJ, 524, 732

Teyssier, R., Chapon, D., & Bournaud, F. 2010, ApJL, 720, L149

Thompson, T. A., Quataert, E., & Murray, N. 2005, ApJ, 630, 167

Toomre, A. 1964, ApJ, 139, 1217

—. 1969, ApJ, 158, 899

—. 1977, ARA&A, 15, 437

Toomre, A. 1981, in Structure and Evolution of Normal Galaxies (Cambridge University Press, NY), ed. S. M. Fall & D. Lynden-Bell, 111–136

Tremaine, S. 2001, AJ, 121, 1776

Volonteri, M., Callegari, S., Colpi, M., Dotti, M., & Mayer, L. 2008, Mem. Soc. Astron. Italiana, 79, 1231

Wada, K. 1994, PASJ: Publications of the Astronomical Society of Japan (ISSN 0004-6264), 46, 165

Weinberg, M. D., & Katz, N. 2007, MNRAS, 375, 460

Wise, J. H., Turk, M. J., & Abel, T. 2008, ApJ, 682, 745

Younger, J. D., Hopkins, P. F., Cox, T. J., & Hernquist, L. 2008, ApJ, 686, 815